

OTDK-dolgozat

2023

BUDAPESTI CORVINUS EGYETEM
INFORMATIKAI INTÉZET



**OSZTÁLYZATELEMZÉSI KERETRENDSZER KIALAKÍTÁSA A
FELSŐOKTATÁSI KÉPZÉSEK STRUKTÚRÁJÁNAK FEJLESZTÉSÉRE**

**DEVELOPING A GRADING ANALYTICS FRAMEWORK TO IMPROVE
THE STRUCTURE OF HIGHER EDUCATION**

Bartha Anna és Galambos Ágnes

Konzulensek: Dr. Vas Réka Franciska és Kovács László

Kézirat lezárásának dátuma: 2021.04.11.

TARTALOMJEGYZÉK

1. BEVEZETŐ (BARTHA ANNA, GALAMBOS ÁGNES)	1
2. TANULÁSELEMZÉS (BARTHA ANNA)	2
2.1. A TANULÁSELEMZÉST ÖSZTÖNZŐ TÉNYEZŐK	3
2.1.1. Növekvő adatmennyiség	3
2.1.2. Online tanulás	3
2.1.3. Politikai érdekek	4
2.2. TANULÁSELEMZÉSHEZ KAPCSOLÓDÓ TERÜLETEK	4
2.2.1. Üzleti intelligencia	4
2.2.2. Webanalitika	5
2.2.3. Adatbányászat	5
2.3. A TANULÁSELEMZÉS, AZ OKTATÁSI ADATBÁNYÁSZAT ÉS AZ AKADÉMIAI ELEMZÉS	6
3. A TANULÁSELEMZÉS FOLYAMATA (BARTHA ANNA)	8
3.1. AZ IGÉNY MEGHATÁROZÁSA	8
3.2. ADATOK RÖGZÍTÉSE	9
3.3. AZ ADATOK ELŐKÉSZÍTÉSE	9
3.4. ELEMZÉS ÉS ÉRTELMEZÉS	10
3.4.1. Statisztikai modellezés	10
3.4.2. Klaszterelemzés	10
3.4.3. Tartalomelemzés	10
3.4.4. Szociális háló elemzése	10
3.5. BETEKINTÉS ÉS ÉRTÉKTEREMTÉS	11
3.6. VISSZAJELZÉS	11
4. OSZTÁLYZATELEMZÉS FOGALMA ÉS MÓDSZEREI (GALAMBOS ÁGNES)	12
4.1. A RENDELKEZÉSRE ÁLLÓ ADATOK ÉS A TERVEZETT VIZSGÁLATOK	13
4.1.1. A rendelkezésre álló adatok	13
4.1.2. A sorrendi skálán mért adatok	13
4.1.3. Az adatok megismeréséhez szükséges statisztikai eszközök	14
4.1.3.1. Helyzetmutatók	14
4.1.3.2. Alakmutatók	15
4.1.4. A változók közötti kapcsolatok vizsgálata	15
4.1.4.1. Korreláció	15

4.1.4.2.	Regresszió	16
5.	OSZTÁLYZATELEMZÉS A GAZDASÁGINFORMATIKA KÉPZÉSBEN (GALAMBOS ÁGNES)	18
5.1.	A RENDELKEZÉSRE ÁLLÓ ADATOK BEMUTATÁSA	18
6.	AZ ADATOK ELŐKÉSZÍTÉSE (BARTHA ANNA).....	19
6.1.	A KAPOTT ATTRIBÚTUMOK	19
6.2.	ANOMÁLIÁK KEZELÉSE.....	20
6.3.	ADATSTRUKTÚRA KIALAKÍTÁSA – AGGREGÁLÁS.....	23
6.3.1.	Aggregálás a hallgatók mentén	23
6.3.2.	Aggregálás az évfolyamok mentén	23
6.3.3.	A végleges adatstruktúra	24
6.3.4.	Hiányzó adatok kezelése az aggregált adatbázisban	24
7.	A KIDOLGOZOTT OSZTÁLYZATELEMZÉSI MEGOLDÁSOK BEMUTATÁSA.	26
7.1.	LEÍRÓ STATISZTIKA (BARTHA ANNA)	26
7.1.1.	Hiányzó és szelektált adatok	26
7.1.2.	A változók vizsgálata	28
7.1.3.	A leíró statisztikai eredmények ismertetése az IRF és ÜI kurzusokon keresztül.....	33
7.2.	METRIKA KIALAKÍTÁSA (GALAMBOS ÁGNES)	34
7.2.1.	A metrika szükségessége.....	34
7.2.2.	Elméleti kérdések a kialakítandó metrikával kapcsolatban.....	34
7.2.3.	A metrika kialakításához alkalmazott eszközök	36
7.2.3.1.	Érzékenységvizsgálat	36
7.2.3.2.	Táblázatok az értékkészlet vizsgálatára	37
7.2.3.3.	Oszlopdiagramok és gyakorisági táblák az értékkészlet és az eloszlások vizsgálatára.....	38
7.2.4.	Gyakorlati jegyes tárgyak.....	39
7.2.5.	A három metrika.....	39
7.2.5.1.	Az a) modell.....	40
7.2.5.2.	A b) modell	41
7.2.5.3.	A c) modell	43
7.2.6.	A három metrika vizsgálata az IRF és ÜI kurzusok eredményein	45
7.3.	RANGKORRELÁCIÓ (GALAMBOS ÁGNES).....	47
7.3.1.	Rangkorrelációs mátrix kialakítása	47

7.3.2.	Rangkorreláció vizsgálata a három adatbázisra	48
7.3.2.1.	Az Összes adatbázis:	48
7.3.2.2.	Az Összes null adatbázis.....	49
7.3.2.3.	A Szűrt adatbázis	49
7.3.3.	A rangkorrelációs mátrixok összehasonlítása a három adatbázis esetében.....	50
7.3.3.1.	Összes és Szűrt közötti eltérések:	50
7.3.3.2.	Összes null és Szűrt közötti eltérések:	51
7.3.3.3.	Összes és Összes null közötti eltérések:	52
7.3.4.	A rangkorrelációs eredmények vizsgálata az IRF és ÜI kurzusokon.....	53
7.3.5.	Döntés a továbbiakban használt adatbázis és metrika mellett.....	54
7.4.	ELŐFELTÉTELEK VIZSGÁLATA (<i>GALAMBOS ÁGNES</i>).....	55
7.4.1.	Előfeltételmátrix.....	55
7.4.2.	Előfeltételes regresszió.....	56
7.5.	LAD REGRESSZIÓ (<i>BARTHA ANNA</i>).....	57
7.5.1.	A tantárgyak közötti kapcsolat vizsgálatának eszközei	58
7.5.2.	A változók közötti kapcsolatok feltárása	61
7.5.3.	Az IRF és ÜI kurzusok eredményeinek bemutatása	66
7.6.	KÖVETKEZTETÉSEK ÉS JAVASLATOK (<i>GALAMBOS ÁGNES</i>).....	69
8.	ÖSSZEFOGLALÁS (<i>BARTHA ANNA, GALAMBOS ÁGNES</i>)	73
	IRODALOMJEGYZÉK.....	77
	MELLÉKLETEK.....	I
1.	MELLÉKLET: VIZSGÁLT TANTÁRGYAK ÉS RÖVIDÍTÉSEIK	I
2.	MELLÉKLET: INFORMATIKAI RENDSZEREK FEJLESZTÉSE TANTÁRGYI ADATLAP A 2014/2015 TANÉVBŐL	I
3.	MELLÉKLET: ÜZLETI INTELLIGENCIA TANTÁRGYI ADATLAP A 2014/2015 TANÉVBŐL	IV
4.	MELLÉKLET: GAZDASÁGINFORMATIKUS (BSC) ALAPKÉPZÉSI SZAK 2012/2013. OPERATÍV TANTERVE	VII

TÁBLÁZAT- ÉS ÁBRAJEGYZÉK

1. táblázat: A tantárgyakhoz tartozó táblázatok kezdeti állapota	19
2. táblázat: A tisztított adattáblák felépítése.....	20
3. táblázat: Operatív tanterven kívüli tárgyfelvevételek száma és aránya az összes adathoz képest.	22
4. táblázat: Példa a végleges adatstruktúrára.....	24
5. táblázat: Hiányzó adatok aránya	25
6. táblázat: Rendelkezésre álló és kiszűrt adatok száma	27
7. táblázat: Elemszám tantárgyi bontásban	27
8. táblázat: Tárgyfelvevételek számának vizsgálata	31
9. táblázat: A vizsgázások számának vizsgálata	32
10. táblázat: Érzékenységvizsgálat a metrika értékeire közepes érdemjegy esetén	37
11. táblázat: Értékkészlet vizsgálata jó érdemjegyre vonatkozóan	38
12. táblázat: Az a) metrika értékkészlete gyakorlati és vizsgajegyű tárgyak esetében.....	41
13. táblázat: A b) metrika értékeinek alakulása gyakorlati és vizsgajegyű tárgyak esetében.....	42
14. táblázat: Kitevőket tartalmazó táblázat a c) modell esetében	43
15. táblázat: A c) metrika értékkészlete elégtelen és közepes érdemjegyre.....	44
16. táblázat: IRF-ből elért eredmények és metrika értékeik.....	46
17. táblázat: ÜI-ből elért eredmények és metrika értékeik.....	46
18. táblázat: Rangkorrelációs mátrix a c) metrika értékeire az Összes null adatbázison	47
19. táblázat: Rangkorrelációs mátrix a c) metrika értékeire az Összes adatbázison	48
20. táblázat: Rangkorrelációs mátrix a c) metrika értékeire a Szűrt adatbázison	50
21. táblázat: Összes és Szűrt rangkorrelációs mátrixok különbsége a c) metrika értékein	51
22. táblázat: Összes null és Szűrt rangkorrelációs mátrixok különbsége a c) metrika értékein.....	52
23. táblázat: Összes és Összes null rangkorrelációs mátrixok különbsége a c) metrika értékein ..	52
24. táblázat: Előfeltétel mátrix	55
25. táblázat: Értékkészlet elemszáma tantárgyanként	59
26. táblázat: Vizsgált változók a LAD regresszióban	62
27. táblázat: Dummy változó kialakítása az évfolyamokra.....	62
28. táblázat: LAD regressziós eredmények.....	65
29. táblázat: Rangkorrelációs mátrix a regressziós modellek alapján szignifikáns kapcsolatokkal	70

1. ábra: A tanuláselemzés folyamata.....	8
2. ábra: Hiányzó adatok arányának változása	28
3. ábra: Nem szerinti megoszlás.....	29
4. ábra: Érdemjegyek eloszlása	30
5. ábra: A vizsgázások számának átlaga és a tárgyak típusa.....	32
6. ábra: Metrika értékek gyakorisági eloszlása oszlopdiagramon.....	39
7. ábra: LAD regresszió értékei.....	57
8. ábra: LAD regresszió az R-ben	60
9. ábra: RStudio – a program futtatása.....	63
10. ábra: LAD_stepAIC output	64
11. ábra: Előfeltétel-rendszer javaslat	71
12. ábra: Osztályzatelemzési keretrendszer folyamatábrája	73

1. BEVEZETŐ

A digitalizáció az élet egyre több területén jelenik meg, amely a digitális formában eltárolt adatok mennyiségének növekedésével jár együtt. Az üzleti életben hamar felfedezték, hogy az adatokból kinyerhető információ értéket teremt, és komoly erőforrást jelent a mindennapi működés során. Ennek megfelelően hamar kialakult az igény az adatok hatékony elemzésére, aminek eredményeképp új elemzési módszerek és tudományágak alakultak ki. Ilyen például az üzleti intelligencia, az adatbányászat és a gépi tanulás. Habár ezek a területek sok szempontból eltérnek egymástól, közös céljuk az adatalapú döntéshozatal megteremtése és támogatása.

Nemcsak az üzleti életben, de az oktatási intézmények folyamataiban is egyre fontosabbá válik az adatok megfelelő hasznosítása. A felsőoktatási intézmények meghatározó feladatává vált a tanítási és tanulási folyamatok nyomon követése annak érdekében, hogy a tanulói teljesítményméréseknek, valamint a nemzetközi akkreditációs központok elvárásainak eleget tudjanak tenni. Emellett kialakult az igény a rendelkezésre álló adatok hatékony elemzése iránt is, melynek célja, hogy feltárja az oktatás területén a fejlesztési lehetőségeket. A felsőoktatásban az adatokra épülő döntéshozatal három fő területre építkezik, a tanulóelemzésre, az oktatási adatbányászatra és az akadémiai elemzésre. Habár egyre több külföldi egyetem életében mindennapos ezen módszerek alkalmazása, hazánkban még csak kezdetleges ezeknek a területeknek a megjelenése.

Kutatásunk során arra a kérdésre kerestük a választ, hogy a magyar felsőoktatásban használt Neptun rendszerben tárolt adatokra építve ki lehet-e alakítani olyan tanulóelemzési folyamatot, melynek célja az adatvezérelt döntéshozatal támogatása. Kíváncsiak vagyunk arra, hogy ezekből az adatokból milyen információkat tudunk kinyerni az egyes kurzusok, és átfogóan egy-egy szak kapcsán. Vajon a rendelkezésre álló (Neptun) adatok által nyújtott korlátos lehetőségek elegendőek ahhoz, hogy módszerek olyan halmazát dolgozzuk ki, melyek támogatják az egyetemi képzések minőségének javítását, a képzések struktúrájának fejlesztését?

A Budapesti Corvinus Egyetem Információrendszerek Tanszéke rendelkezésünkre bocsátotta a Gazdaságinformatikus BSc képzésén oktatott informatikai-matematikai tárgyakhoz kapcsolódó Neptun adatokat, három – tanulmányait befejezett – évfolyam kapcsán. A célunk tehát az volt, hogy feltárjuk ezen kurzusok hallgatói teljesítményei közti összefüggéseket, és meghatározzunk olyan módszereket, amelyek segítségével átfogó képet nyerhetünk a képzési struktúráról.

A vizsgálat során kialakított módszerek elsősorban a tantervi háló tervezését támogathatják, ezzel elsődlegesen az oktatók és szakfelelősök munkáját segítve. Az elemzések során kapott eredményekből láthatjuk, hogy mely tárgyak között van szignifikáns kapcsolat, illetve a korábbi félévek tárgyaiból elért eredmények milyen mértékben befolyásolják a későbbi hallgatói teljesítmény alakulását. Ezek alapján könnyen kialakítható olyan előkövetelmény-rendszer, amely az órákon oktatott anyagok és megszerzett készségek tényleges egymásra épülését támaszthatja alá.

A dolgozatunk négy részből áll. Első körben a tanulóelemzés, és az ehhez kapcsolódó osztályzatelemzés területét mutatjuk be részletesen, majd az adatok előkészítésének folyamatát ismertetjük. Következő lépésben az adatokon végzett elemzések eredményeit foglaljuk össze. Ennek során kitérünk többek között egy új teljesítmény metrika kialakítására, valamint a LAD regresszió keretében alkalmazott változószelekciós algoritmusra is. Végül pedig bemutatjuk, hogy az összeállított elemzési módszerek halmaza miként hasznosítható a Gazdaságinformatikus képzés struktúrájának fejlesztésére.

2. TANULÁSELEMZÉS

Az analitika kifejezés napjainkra egyre hangsúlyosabbá válik, és gyakran már leghétköznapiabb tevékenységeink során is találkozhatunk vele. Az elemzés alapú megközelítés legnagyobb mértékben az üzleti területeken van jelen, azonban az oktatás kapcsán is egyre gyakrabban merül fel (Zilvinskis–Willis, 2019). A tanulóelemzés (Learning Analytics, LA) a technológiailag támogatott tanulás (Technology-Enhanced Learning, TEL) gyorsan fejlődő része, de szorosan összekapcsolódik számos más területtel is, mint például az üzleti intelligenciával, a webanalitikával és az adatbányászattal. Mivel módszertanilag ilyen sok tudományterület eszközeiből merít, a kutatók és a gyakorlati szakemberek számos szempontból közelítették meg ezt a területet (Ferguson, 2012).

A 2011-ben tartott első nemzetközi Learning Analytics and Knowledge konferencián az alábbiak szerint határozták meg a tanulóelemzés fogalmát: „A tanulóelemzés a tanulókról és azok kontextusairól szóló adatok mérése, gyűjtése, elemzése és ábrázolása, a tanulás és ennek környezetének megértése és optimalizálása céljából.” (Ferguson, 2012) Ezt a definíciót később a Society for Learning Analytics Research (SoLAR) is átvette, annak érdekében, hogy egységes megközelítése legyen ennek a területnek (Shum–Ferguson, 2012).

Ahhoz, hogy jobban átlássuk a tanulóelemzés folyamatát, és megértsük kialakulásának okait, érdemes megvizsgálni, hogy mely tényezők teszik lehetővé és ösztönzik a terület fejlődését.

2.1. A TANULÁSELEMZÉST ÖSZTÖNZŐ TÉNYEZŐK

A tanuláselemzés fejlődésére alapvetően három terület gyakorolt nagy hatást: a növekvő adatmennyiség, az online oktatás és a politikai érdekek. A továbbiakban ezeket részletesebben is megvizsgáljuk (Ferguson, 2012).

2.1.1. Növekvő adatmennyiség

A Big Data, mint fogalom a nagy mennyiségű (Volumen), nagy sebességgel változó (Velocity) és nagyon változatos (Variety) adatokat jelenti, gyakran utalnak rá 3V-ként. A terület legnagyobb kihívása a rendszerekbe beérkező adatok nagyon rövid idő alatt történő rögzítése és feldolgozása (Vas, 2021). A virtuális tanulási környezetek (Virtual Learning Environment, VLE), vagy más néven tanulásirányítási rendszerek (Learning Management System, LMS), mint például a Moodle, széles körű bevezetésének következtében az oktatási intézmények is egyre nagyobb adathalmazokkal tudnak foglalkozni, ám ezek tárolását, feldolgozását és megfelelő alkalmazását is meg kell oldaniuk (Mining, 2012). Habár az oktatásban még nem beszélhetünk Big Data-ról, de ezek a rendszerek fokozatosan egyre nagyobb mennyiségű interakciós és személyes adatot, rendszerinformációt és akadémiai információt halmoznak fel (Ferguson, 2012).

Annak ellenére, hogy a hallgatói tevékenység nyomon követése sok esetben általános szoftverfunkcióként szerepel a tanulásirányítási rendszerekben, ezekhez nem kapcsolódnak használható elemzési módszerek. Ezenfelül az adatok kibontására, összesítésére és megjelenítésére vonatkozó funkciók is gyakran hiányosak, vagy nem is léteznek. A másik jelentős problémát az okozza, hogy sokszor egy-egy intézmény többféle rendszert is használ, így az adatok különböző forrásrendszerekből érhetők el, eltérő szabványokkal, struktúrával és hozzáférési szintekkel (Elias, 2011).

Az első ösztönző tényező tehát egy technikai kihívás: A tanulásirányítási rendszerekből hogyan gyűjtsük és kezeljük egységesen az adatokat, majd ezeket a tanulási adatokat felhasználva hogyan teremtsünk értéket (Ferguson, 2012)?

2.1.2. Online tanulás

Az oktatásban a növekvő adatmennyiség megjelenésének előfeltétele az online tanulás elterjedése volt. Ez amellet, hogy számos előnnyel járt, korábban nem ismert problémákat is felvetett (Ferguson 2012). A hallgatók elszigetelten érezhetik magukat, különböző technikai problémákat tapasztalhatnak, és így a motivációjukat is könnyen elveszíthetik. A tanárok ugyanezen nehézségek mel-

lett azzal is szembesülhetnek, hogy nehezebben kapnak képet arról, hogy a hallgatók milyen szinten állnak, így könnyen teljesítményromláshoz vezethet az online tanulási és tanítási folyamat. A tanulóelemzés kialakulását tehát az is elősegítette, hogy az online térben történő eseményeket nyomon tudjuk követni. Ezzel párhuzamosan pedig felmerült az igény, hogy ezeknek az adatoknak a segítségével optimalizáljuk a folyamatokat hallgatói és oktatói szempontok alapján (Florence–Abdou, 2016).

2.1.3. Politikai érdekek

Egyre több országban nő az igény az iránt, hogy az oktatási intézményekben a lehető legtöbb területen jelen legyen a teljesítmény mérése, nyomon követése és javítása. Az adatok, és az ezekre épülő tudás értéket teremt, valamint hatalmat is jelent a birtokosának. Az oktatást tekintve pedig nemzeti és nemzetközi cél a folyamatos javulás, és olyan magas színvonalú oktatási rendszer kiépítése, amelyet az adatok segítségével monitorozni, és optimalizálni tudunk. Ennek megfelelően a politikai érdekeket tekintjük a harmadik hajtóerőnek (Shum–Luckin, 2019).

Ez a három mozgatórugó három különböző csoport érdekeire hívja fel a figyelmet: kormányok, oktatási intézmények, illetve tanárok és tanulók. Habár ezek az érdekek sok esetben átfedik egymást, a szükséges analitikai módszerek típusa és megközelítése eltérhet az egyes esetekben. A közös pont az, hogy mind a három esetben az oktatási folyamatok optimalizálása a cél, azonban az érdekcsoportok nagyban befolyásolják, hogy az elemzők hogyan fogalmazzák meg a konkrét problémát, gyűjtik az adatokat, értelmezik az eredményeket, és végül miként implementálják a megoldásokat (Ferguson, 2012).

2.2. TANULÁSELEMZÉSHEZ KAPCSOLÓDÓ TERÜLETEK

A tanulóelemzés kialakulását tehát több tényező is ösztönzi, így ennek az oktatásba való implementálása már elkezdődött és napjainkban is zajlik. Az LA azonban eltér a teljesen új technológiai és módszertani megoldásoktól, mivel más területekkel szorosan összekapcsolódva fejlődik. Ezek a korábban már említett üzleti intelligencia, webanalitika és adatbányászat (Shum–Ferguson, 2012).

2.2.1. Üzleti intelligencia

„Az üzleti intelligencia azon eljárások, módszerek és megoldások összessége, melyek a vállalati adatok másodlagos hasznosítása, felhasználása révén támogatják a döntéshozatalt.” (Fajszi–Cser–Fehér, 2010, p. 39.) A tanulóelemzés kifejezés legkorábbi említése is ehhez a fogalomhoz köthető,

ez akkorra tehető, amikor az üzleti intelligencia keretein belül az e-learning termékek és szolgáltatások területét kezdték vizsgálni (Shum–Ferguson, 2012).

2.2.2. Webanalitika

A 21. században a Web 2.0 megjelenése új lehetőségeket nyitott a webes tartalmak különböző forrásokból történő összegyűjtésére, feldolgozására és az eredmények programok közötti közvetítésére (Shum–Ferguson, 2012). A webkettő azonban sokkal több mint pusztán technológia. Ez olyan internetes szolgáltatások összessége, ahol a felhasználók maguk is állíthatnak elő tartalmat, megoszthatják egymás információit. Ebben az esetben a szervergazda feladata mindössze a keretrendszer biztosítása, a tartalmat maguk a felhasználók hozzák létre, osztják meg vagy véleményezik (Hicks–Graber, 2010). Ennek során nagy mennyiségű adat keletkezik, melyek felhasználására és az azokból való értékteremtés céljából alakult ki a webanalitika. Ez a weboldalak forgalmának mérését teszi lehetővé, valamint az oldal teljesítményének javításában is segítséget nyújt (Zheng–Peltsverger, 2015). A virtuális tanulási környezetek (Virtual Learning Environments, VLE) gyors elterjedése is a Web 2.0 megjelenéséhez fűződik, hiszen ezeken a platformokon a hallgatók nemcsak aktív felhasználók, de tartalomgyártók is lehetnek. Az Egyesült Királyságban 1994-ben a felsőoktatási intézmények 7%-a használt ilyen rendszert, 2001-ben ez az arány már 40%-ra, 2003-ra pedig több mint 85%-ra nőtt. A VLE-k alkalmazása nagymennyiségű, hallgatói tevékenységhez fűződő adat keletkezését is jelenti. Ezek kiaknázására és felhasználására webanalitikai módszerek is alkalmazásra kerültek. (Shum–Ferguson, 2012).

2.2.3. Adatbányászat

„Az adatbányászat az adatbázisokból olyan implicit és rejtett információkat, összefüggéseket, szabályszerűségeket, mintázatokat nyer ki, amelyek a gyakorlatban is jól hasznosíthatók.” (Fajszi et al., 2010, p. 22.). Ez a terület a számítástechnika számos kutatási területét egyesíti, ideértve többek között a mesterséges neurális hálózatot, a Bayes tanulást, a döntési fa felépítését, a példányalapú tanulást és a statisztikai algoritmusokat is. Az adatbányászat alkalmazásának igénye a virtuális tanulási környezetek és az oktatásirányítási rendszerek elterjedésével jelent meg az oktatásban. Ennek oka, hogy az ezen rendszerekben felhalmozódó, nagymennyiségű adatokból az érintettek tudást szerettek volna kinyerni (Baradwaj–Pal, 2011).

2.3. A TANULÁSELEMZÉS, AZ OKTATÁSI ADATBÁNYÁSZAT ÉS AZ AKADÉMIAI ELEMZÉS

Láthattuk tehát, hogy a különböző érdekek mentén kialakult egy igény az oktatásban megjelenő adatok hasznosítására. A megoldást pedig nem egy teljesen új módszer megjelenése jelentette, hanem az üzleti intelligenciára, a webanalitikára és az adatbányászatra építve alakult ki a tanulás-elemzés és az oktatási adatbányászat (Educational Data Mining, EDM), majd később ebből a kétből vált ki egy harmadik terület, az akadémiai elemzés (Academic Analytics, AA) (Charitopoulos–Rangoussi–Koulouriotis, 2020). A következőkben szeretnénk rávilágítani ennek a három területnek az összefüggéseire és eltéréseire, hogy meg tudjuk határozni, hogy a kutatásunk során megfogalmazott hipotézist milyen irányból érdemes megközelíteni.

Az EDM új algoritmusok és modellek kialakításával és alkalmazásával foglalkozik annak érdekében, hogy az oktatási adatok nagy halmazában található mintázatokot feltárja, amelyeket egyébként nehéz vagy sok esetben szinte lehetetlen elemezni a meglévő nagy adatmennyiség miatt. Az oktatási adatbányászat tehát alapvetően összegyűjti, és elemzi az e-learning platformok adatbázisaiban tárolt adatokat (Mining, 2012).

Az LA ezzel szemben az adatok által vezérelt döntéshozatalt jelenti az oktatás területén. Ez inkább az ismert prediktív modellekre épít, és olyan adatelemzési módszerek és eszközök alkalmazását jelenti, amelyek az oktatással kapcsolatos kérdések megválaszolását támogatják. A tanulás-elemzés alapvetően a tanulóra összpontosít, és a különböző rendszerekből összegyűjtött adatok segítségével igyekszik a hallgatói sikereket támogatni, és az esetlegesen felmerülő problémák esetén figyelmeztetni az érintetteket (Charitopoulos et. al., 2020).

Az EDM és a LA területek tehát szorosan kapcsolódnak egymáshoz, és sok esetben együttesen is tárgyalják őket. A látszólagos hasonlóságok ellenére azonban több dologban is felfedezhetőek különbségek. A feltárási és adaptációs módszerek, illetve maga a megközelítési szemlélet is eltérő. Ez utóbbi lehet redukcionista (a rendszer elemeit részekre bontja) vagy holisztikus (az egész az több mint a részek összege) (Charitopoulos et. al., 2020). Ennek megfelelően az alábbi különbségeket lehet beazonosítani:

Az EDM-hez fűződő módszerek és eszközök összegyűjtik és csoportosítják, vagy újra strukturálják az adatokat annak érdekében, hogy felismerjék az ezekben rejlő mintákat és kapcsolatokat. Ezek a lépések napjainkra már nagyrészt automatizáltan történnek. Ezzel szemben az LA a fel-

tárt minták és kapcsolatok magasabb szintű értékelésére törekszik, melynek célja, hogy az oktatásban felmerülő kutató jellegű kérdésekre választ adjon. Az ilyen feladatok nagymértékben az emberi megítélésen alapszanak (Mining, 2012).

Az EDM továbbá redukcionista megközelítést alkalmaz, és egy konkrét oktatási esemény problémaival vagy összetevőivel foglalkozik. Ezzel szemben az LA a holisztikus megközelítést preferálja, és megkísérli az oktatási folyamat egészét vizsgálni, így kialakítva egy átfogó nézőpontot (Charitopoulos et. al., 2020).

A harmadik terület, az akadémiai elemzés a tanulóelemzés és az oktatási adatbányászat keveréke. Az Academic Analytics a nagy adatkészletek és a prediktív modellezés kombinációjának tekinthető, melynek célja a döntéshozatal javítása. A felsőoktatásban megjelenő fokozott verseny, az akkreditációs folyamatok, az értékelés és a szabályozás egyre nagyobb nyomást gyakorolnak az intézményekre. Ezek hatására a felsőoktatásban is kialakult az igény az adatvezérelt döntéshozatal iránt (McNaughton–Rao–Mansingh, 2017). Ennek eszköze lett az akadémiai elemzés, mely ennek megfelelően elsősorban az operatív folyamatokat támogatja, és az elsődleges érintettjei a menedzsment és a vezetés (Scapin, 2015).

Nem véletlen, hogy pont három terület alakult ki, hiszen az 2.1. fejezetben a tanulóelemzés ösztönzői kapcsán szintén három tényezőt ismertettünk. Ezek szorosan összefüggnek, hiszen a felmerült igényekre válaszul alakult ki a tanulóelemzés, az oktatási adatbányászat és az akadémiai elemzés (Ferguson, 2012):

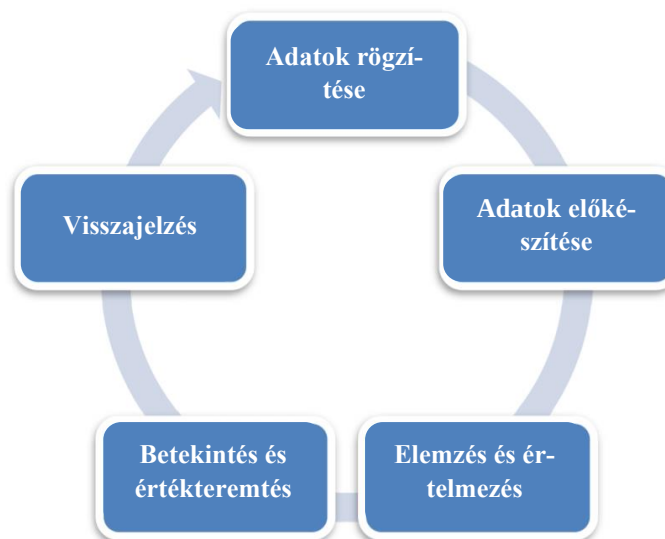
- Az oktatási adatbányászat a technikai kihívásokra összpontosított, és a növekvő adatmennyiség kapcsán felmerülő problémákra reagál: Hogyan vonhatunk ki értéket ezekből a nagy méretű, tanúlással kapcsolatos adathalmazokból?
- A tanulóelemzés az oktatási kihívásokra fókuszál: Hogyan optimalizálhatjuk az online tanulási módszerek lehetőségeit?
- Az akadémiai elemzés a politikai és gazdasági kihívásokra összpontosított: Hogyan javíthatnánk jelentősen a tanulási lehetőségeket és az oktatási eredményeket nemzeti vagy nemzetközi szinten?

Mivel kutatásunk célja a Neptun tanulásirányítási rendszerből kinyerhető adatok hasznosságának vizsgálata, és annak feltárása, hogy ezek alapján miként lehetne a felsőoktatási képzések struktúráját javítani, így ennek elsődleges érdekeltjei a tanárok, a szakfelelősök és a szakfejlesztők.

Továbbá ez egy olyan probléma, amely az oktatási folyamat egészét vizsgálja, így ennek megfelelően a hipotézisünk vizsgálatára a tanuláselemzési módszer alkalmazása mellett döntöttünk.

3. A TANULÁSELEMZÉS FOLYAMATA

Ahhoz, hogy az oktatásban megfelelő módon alkalmazni tudjuk a tanuláselemzés adta lehetőségeket, fontos megismerni és tisztában lenni ennek folyamatával. Ahogy azt az 1. ábra mutatja, a tanuláselemzés egymással összefüggő lépések vagy szakaszok sorozataként fogható fel: 1) a tanulók által generált adatok rögzítése; 2) az adatok feldolgozása elemzés céljából; 3) elemzés során használt statisztikai és gépi tanulási módszerek használata; 4) a kapott adatokra épülő eredmények felhasználása a döntéshozatali folyamat elősegítésére; 5) tekintettel arra, hogy a tanuláselemzés iteratív folyamat, a visszacsatolás segíthet a módszerek fejlesztésében és finomításában (Saqr, 2018).



1. ábra: A tanuláselemzés folyamata

Forrás: Saqr, 2018, p. 39.

3.1. AZ IGÉNY MEGHATÁROZÁSA

Mielőtt azonban részletesen kifejtjük ennek a ciklusnak az öt lépését, fontos látni, hogy szükséges egy nulladik pont azonosítása is. Ahhoz, hogy egy tanuláselemzési folyamatot sikeresen implementálni tudjunk, elengedhetetlen, hogy legyen egy meghatározott igény arra vonatkozóan, hogy mit szeretnénk elérni, a tanuláselemzés megvalósításának célja definiálva legyen. Ezenfelül a stakeholdereket és az érintetteket is be kell tudni azonosítani, akik azonban nem minden esetben

egyeznek meg azokkal, akik számára majd a legtöbb hasznot jelentik az eredmények. A korábbiakban már láthattuk, hogy a tanuláselemzés az adatok által vezérelt döntéshozatalt segíti elő az oktatás területén, és alapvetően a tanárok és a hallgatók munkáját segíti, így sok esetben az oktatók lesznek a stakeholderek egy-egy LA igény esetében (Tsai–Poquet–Gašević–Dawson–Pardo, 2019).

3.2. ADATOK RÖGZÍTÉSE

Amennyiben a fenti kérdéseket tisztáztuk, akkor kezdődhet el a tanuláselemzés tényleges folyamata, mely az adatok rögzítésével indul. Minden tevékenység, amelyet egy hallgató online végez, rögzíthető és felhasználható a tanuláselemzés céljából, azonban mindig figyelembe kell venni az etikai és adatvédelmi normákat is (Dyckhoff–Zielke–Bültmann–Chatti–Schroeder, 2012). Az összegyűjthető adatok köre folyamatosan bővül a továbbfejlesztett tanulási technológiák és az érdekelt felek tudatosságának növekedése miatt. A tanulásirányítási rendszerek bevezetése lehetővé tette az olyan adatok gyűjtését, mint a bejelentkezések gyakorisága, az oldalmegtekintések száma, a tanulási források találatai és a feladatok elvégzésére fordított idő nagysága. Ezenfelül egyéb rendszerekből összegyűjthetőek regisztrációs és demográfiai értékek is, mint például az életkor, a nem, a nyelvtudás, a társadalmi-gazdasági helyzet, az ösztöndíjak és a lakóhely (Zilvinskis–Willis, 2019).

3.3. AZ ADATOK ELŐKÉSZÍTÉSE

Ebben a szakaszban több folyamatot is végre kell hajtani annak érdekében, hogy az adatokat megfelelően előkészítsük az elemzésre. Mivel az adatokat sok esetben különböző forrásrendszerekből kell összegyűjteni, de végül egy adatbázisba szeretnénk őket osszerendezni, ezért nagyon fontos az adatelőkészítés lépése. Ennek során feltárásra kerülnek a különböző adatminőségi hibák, mint például az érvényességi vagy mezők közötti hibák, a pontosság vagy a konzisztencia hiánya. Ezek javítása után következik az adatátalakítás lépése, amikor az elemzéshez szükséges módosításokat végezzük el az adatokon. Végül az így előállt értékeket aggregálni kell annak érdekében, hogy az esetlegesen több forrásból származó adatokat összekapcsoljuk, és egy egységes struktúrával tudjunk a továbbiakban dolgozni (Fajszi et. al., 2010).

3.4. ELEMZÉS ÉS ÉRTELMEZÉS

A folyamat következő lépése az adatelemzési módszerek meghatározása és alkalmazása. A következő pár bekezdésben a tanuláselemzéshez leginkább kötődő módszereket mutatjuk be.

3.4.1. Statisztikai modellezés

A statisztikai modellezés elsősorban a tanulásirányítási rendszerek által rögzített információkat, valamint a hallgatók demográfiai és regisztrációs adatait használja fel. Leggyakrabban kétféle módszer kerül felhasználásra. Az első típus, a magyarázó modellezés egy előre definiált hipotézis ellenőrzéséhez keres oksági kapcsolatokat a változók között. A második típus a prediktív modellezés, amelynek fő célja a rendelkezésre álló adatok felhasználása egy jövőbeli esemény vagy kimenetel korai előrejelzéséhez (Saqr, 2018).

3.4.2. Klaszterelemzés

A klaszterelemzés célja olyan adatpontok feltárása, amelyek mentén egy teljes adathalmazt kategóriákra lehet bontani úgy, hogy a hasonló tulajdonsággal rendelkező egyedek egy csoportba kerülnek. Az oktatásban például csoportosíthatjuk a hallgatókat a tanulási nehézségeik és interakciós szokásaik alapján, majd ennek megfelelően személyre szabott ajánlások készíthetők a hallgatók részhez számára. Ez a módszer szorosabban kapcsolódik az oktatási adatbányászathoz, de bizonyos esetekben a tanulóelemzés során is alkalmazzák (Mining, 2012).

3.4.3. Tartalomelemzés

A tartalomelemzés középpontjában az olyan tartalmak állnak, mint például a tanuláshoz kapcsolódó könyvek, segédanyagok és webes források, valamint a vitafórumok, értékelések, feladatok és a társas interakciók tartalma. A tartalomelemzés egy automatizált módszer, amely a digitális tanulási tartalom különböző formáit vizsgálja, értékeli, szűri és vizualizálja, a tanulási tevékenységek megértése és az oktatási tevékenység javítása céljából (Saqr, 2018).

3.4.4. Szociális háló elemzése

A szociális háló elemzése az adatelemzés olyan területe, amely a társadalmi struktúrák és kapcsolatok feltérképezésével és az emberek közötti interakciókkal foglalkozik. A matematika területéről a gráfelméletet, míg a szociológia vonatkozásában az interperszonális és közösségi kapcsolatok vizsgálatára használt eszközöket egyesíti (Shum–Ferguson, 2012). Az oktatásban is nagy jelentősége van, hiszen ennek a módszernek a segítségével fel lehet térképezni a hallgatók közötti kapcsolatokat. Például az iskolai barátok egy csoportja a baráti kapcsolat által meghatározott hálózatot

alkot, amely irányítatlan gráfként reprezentálható. A szereplőket általában csomópontoknak vagy csúcsoknak nevezik; az élek pedig azok a kapcsolatok, amelyek összekapcsolják a csomópontokat, és így egy társadalmi struktúrát alakítanak ki (Saqr, 2018). A közösségi hálózat elemzése hasznos eszköz többek között az online tanulás vizsgálatához, mivel az interperszonális kapcsolatok fejlesztésére összpontosít úgy, hogy magát a technológiát is a folyamat részeként tekinti. Ennek során olyan beavatkozási pontok azonosíthatók, amelyek segítségével támogatni lehet a hálózat tagjainak tanulási folyamatát úgy, hogy megfelelő kapcsolatokat, erőforrásokat és ötleteket kapcsolunk hozzájuk (Shum–Ferguson, 2012).

3.5. BETEKINTÉS ÉS ÉRTÉKTEREMTÉS

A tanuláselemzés végső célja, hogy az érdekeltek, a hallgatók és a tanárok számára olyan hasznos információt nyújtson, amelynek segítségével az oktatást javítani lehet. Ennek megfelelően az elemzés során kapott eredményeket első körben értelmezni kell, majd meg kell határozni azt is, hogy ezek milyen formában tudnak értéket teremteni az adott érdekelt számára (Fajszi et. al., 2010).

3.6. VISSZAJELZÉS

A tanuláselemzés folyamatának eredménye betekintést ad az oktatásban történő eseményekre, és segítséget nyújt annak optimalizálásához. Azonban miután az új módszereket elkezdjük alkalmazni, szükséges valamiféle visszacsatolás is, hogy láthatóak legyenek az eredmények és esetlegesen újabb elemzéseket tudjunk végezni. Ezenfelül a technológia fejlődésével további lehetőségek alkalmazása válik elérhetővé, illetve ahogy egyre több adatot tudunk összegyűjteni, úgy az elemzési módszerek is egyre pontosabbá válhatnak (Saqr, 2018).

4. OSZTÁLYZATELEMZÉS FOGALMA ÉS MÓDSZEREI

Az osztályzatelemzés (Grading Analytics, GA) a tanuláselemzés egy fajtája, mely a diákok teljesítményét a kapott érdemjegyekből kiindulva vizsgálja. A GA területét kevés tanulmányban tárgyalják, és nincs egy egyértelműen körülhatárolható fogalma. Ennek következtében az LA mellett gyakran sorolják az AA és az EDM területéhez is, az elemzésekkel elérni kívánt céloktól függően. Az osztályzatelemzés alatt azt a folyamatot értjük, melynek során a hallgatók kurzus végén kapott érdemjegyeit és az azok megszerzéséhez vezető utat vizsgáljuk. Az elemzésekhez a hallgatói eredményekhez kapcsolódó hagyományos adatokat használjuk, azaz nincs szükség a különböző tanulásirányítási platformok és virtuális tanulási környezetek által online, automatikusan gyűjtött adatokra. Mivel a felsőoktatási intézmények mindegyike használ a hallgatói adatok tárolására alkalmas felületet (például Neptun rendszer), így az elemzéshez szükséges adatok mindenhol rendelkezésre állnak (Dziuban–Moskal–Cavanagh–Watts, 2012) (Ferguson, 2012) (Zielke–Bernst–Crombach–Becker–Lippel, 2018).

Az osztályzatelemzés során a már meglévő adatok kiértékelése történik. Segítségével múltira és jelenre vonatkozó elemzések készíthetők a hallgatói teljesítménnyel, a vizsgált képzéssel és az egyes tárgyakkal kapcsolatban. Emellett prediktív módszerekkel előrejelezhető a hallgatói teljesítmény is, amelynek következtében a GA folyamat érdekeltjei lehetnek hallgatók, tanárok és egy-egy vizsgált képzés szakfelelősei is. Az itt megfogalmazott igények és azok érdekeltjei alapján soroljuk az osztályzatelemzést az ezekhez megfelelő eszközöket nyújtó LA területéhez. A GA a tanuláselemzési modellek közül a prediktív és a magyarázó modelleket használhatja. A magyarázó modell alkalmazásakor megfogalmazhatunk hipotéziseket például arról, hogy az egyes tárgyak között milyen mértékű kapcsolat van, amihez jó kiindulási alap a már meglévő tantervi háló. Az elemzés végén pedig előállhat a hipotéziseket alátámasztó vagy cáfoló tantervi háló javaslat, amely ezután beépíthető a képzési struktúrába. Ezzel a módszerrel főként a vizsgált szak szakfelelősei, szakfejlesztői, illetve a tárgyak oktatói számára teremtünk értéket. A prediktív modellezés során pedig korábbi eredmények alapján előre jelezhetjük a későbbiekben várható, egyéni hallgatói teljesítményt (Dzubian et al., 2012) (Saqr, 2018) (Zielke et al., 2018).

4.1. A RENDELKEZÉSRE ÁLLÓ ADATOK ÉS A TERVEZETT VIZSGÁLATOK

A következő szakaszban megvizsgáljuk a GA során vizsgált adathalmaz jellemzőit és sajátosságait a kutatásunk példáján keresztül. Ezt követően összefoglaljuk az osztályzatelemzési folyamat elvégzéséhez szükséges statisztikai eszközök és módszerek általános leírását. Ezen elemzések segítségével lehetőség van az egyes tárgyakból elért hallgatói eredmények jellemzőinek feltárására, a különböző tárgyakból szerzett eredmények közötti kapcsolatok feltárására, az előfeltételek hatásának elemzésére, illetve annak vizsgálatára, hogy hogyan és milyen mértékben hatnak egymásra a különböző tárgyakból elért eredmények.

4.1.1. A rendelkezésre álló adatok

Kiinduló adatforrásunk a Budapesti Corvinus Egyetem Neptun rendszere volt. Az Informatikai Intézet az itt gyűjtött adatokat bocsátotta a rendelkezésünkre, amely lehetővé tette a későbbi adatelemzési folyamat elvégzését. Három olyan gazdaságinformatikus évfolyam adatait vizsgáltuk, akik már befejezték tanulmányaikat az egyetemen. Ezek a 2011/2012, 2012/2013 és a 2013/2014-es őszi félévben kezdett évfolyamok voltak. Rendelkezésünkre álltak a hallgatók egyes tárgyakból elért korábbi eredményei, a végső eredményük, illetve a nemük és az, hogy a vizsgált három évfolyam közül melyikbe jártak. Mivel az érdemjegyek önmagukban nem hordoznak elég információt a hallgatók teljesítményével kapcsolatban, ezért az érdemjegyekből és korábbi eredményekből, azaz, hogy a végső érdemjegy eléréséhez milyen út (hány tárgy- és vizsgafelvétel) vezetett, megalkottunk egy új teljesítmény-metrikát. Segítségével nemcsak egy olyan új változót hoztunk létre, amely magában hordozza az összetett hallgatói teljesítményt, hanem a jegyek öt elemű értékkészletét kiszélesítettük, ezzel lehetővé téve olyan statisztikai elemzések használatát, amelyek szűk értékkészleten nem, vagy csak megkötések mellett alkalmazhatók. A teljesítmény-metrika részletes tárgyalására a 7.2. fejezetben kerül sor (Hunyadi–Vita, 2006) (Foster–Diamond–Banton, 2014).

4.1.2. A sorrendi skálán mért adatok

A hallgatói teljesítmény-elemzéshez használt statisztikai eszközök kiválasztásánál figyelembe kellett vennünk, hogy az érdemjegyek, illetve az ezek alapján kialakított új teljesítmény-metrikák is sorrendi skálán értelmezett változók. Azaz bár egymáshoz képest a sorrendjük megkülönbözteti őket, értékeik nem számszerűek. A nem mennyiségi ismérvértékek következtében ezeken a változókon nem végezhetők matematikai alpműveletek, emiatt pedig az alapvető statisztikai mutatók egy részének értelmezése sem lehetséges (Hunyadi–Vita, 2006) (Foster et. al., 2014).

4.1.3. Az adatok megismeréséhez szükséges statisztikai eszközök

Az elemzések megkezdése előtt fontos, hogy megvizsgáljuk a rendelkezésre álló adatainkat, és célszerű elvégeznünk olyan leíró statisztikai elemzéseket, melyek segítségével átfogó képet kapunk ezekről. Ennek során érdemes szemügyre venni az adatok eloszlását, alak- és helyzetmutatóit.

4.1.3.1. Helyzetmutatók

Egy-egy tárgy nehézségéről, teljesíthetőségéről ad fontos információt, ha megvizsgáljuk többek között, hogy mi volt az az érdemjegy, amit a legtöbben elértek, vagy mi volt a legjellemzőbb érték tárgyfelvételek és vizsgák számánál. Ezekkel a vizsgálatokkal az adataink eloszlásának helyzetét jellemezzük egyetlen szám segítségével. Az így kapott értékek a helyzetmutatók: a medián, a módusz és az átlag. Mivel a vizsgált változóink közül az érdemjegy sorrendi skálán mért adat, ezért figyelniük kell arra, hogy így a helyzetmutatók közül az átlag számítása problémás lehet (Hunyadi–Vita, 2006).

A helyzeti középértékek közül mediánt az érdemjegyekre nincs értelme számítanunk. Ennek oka, hogy a medián a sorba rendezett értékek közül a középső adat, így szűk értékkészlet esetén nem, vagy alig lenne eltérés a tárgyak medián eredményei között, nagyrészt a közepes érdemjegy lenne a középső érték. Azonban az újonnan képzett teljesítmény-metrika értékkészlete elég széles ahhoz, hogy medián szempontjából az egyes tantárgyak megkülönböztethetők legyenek, vagyis így a mutató értéke értelmezhetővé vált (Hunyadi–Vita, 2006) (Foster et. al., 2014).

A módusz, azaz a leggyakoribb érték megállapítására nincsenek ilyenfajta megkötéseink, értékét az eredeti érdemjegyekre is érdemes megvizsgálunk. Fontos különbség és akár visszajelzés is lehet egy tárgy szempontjából, ha a legtöbb hallgató elégséges vagy akár jeles érdemjegyet ért el. De ugyanígy meghatározhatjuk a teljesítmény-metrika legjellemzőbb értékét is, ha az összetett hallgatói teljesítményt akarjuk vizsgálni (Hunyadi–Vita, 2006) (Foster et. al., 2014).

A sorrendi skálán mért változóinkon számításokat nem végezhetünk, ezért ezeken számított középértékeket sem tudunk értelmezni statisztikai szempontból. Azonban a mennyiségi ismérveken kívül (tárgyfelvétel, vizsgafelvétel, vizsgázás száma) mégis érdemes lehet megvizsgálunk az egyes tantárgyakból elért érdemjegyek számtani átlagait is. Ennek oka, hogy ez az a statisztikai mutató, ami az érdemjegyekkel kapcsolatban olyan szinten beépült a köztudatba, hogy bár statisztikailag helytelen a használata, mégis hordoz viszonylag könnyen értelmezhető információt egy-egy tárggyal kapcsolatban (Hunyadi–Vita, 2006).

4.1.3.2. Alakmutatók

Az adatok megismerésére irányuló elemzések során a mennyiségi változók esetén érdemes lehet megvizsgálni az eloszlás alakjára vonatkozó aszimmetria és csúcsosság mutatók értékeit. Értékes információkat adhat egy tárgyról az, hogy a hallgatók jellemzőbben inkább jobb vagy rosszabb eredményt értek el, illetve, hogy egy-egy érdemjegyet kiugróan sokan szereztek meg, vagy a hallgatói eredmények egyenletesebben oszlanak el a végső bejegyzések között. Az alakmutatókkal megvizsgálhatjuk, hogy az adataink gyakorisági eloszlása egy ugyanolyan szórású normális eloszláshoz képest hogyan alakul. Az aszimmetria mértékét vizsgáló Pearson-féle ferdeségmutató -1 és 1 közötti értéket vehet fel, attól függően, hogy a normális eloszláshoz képest jobbra vagy balra elnyúló a gyakorisági eloszlás. A csúcsosság mérésére szolgáló α_4 mutató is azt vizsgálja, hogy a megfigyelt változó gyakorisági eloszlása hogyan alakul a normális eloszláshoz képest. Pozitív érték esetén csúcsosabb, negatív érték esetén pedig lapultabb az eloszlás, értékei az aszimmetria mutatóhoz hasonlóan -1 és 1 között mozoghatnak. Az alakmutatók ordinális skálán mért adatok esetében szemléltetésre alkalmasak, hiszen a normális eloszlás, az alakmutatók referencia-eloszlása folytonos, ami sorrendi adatok esetében nem előállítható (Hunyadi–Vita, 2006).

4.1.4. A változók közötti kapcsolatok vizsgálata

Az osztályzatelemzési folyamat célja lehet többek között az egyes tárgyakból elért hallgatói eredmények közötti kapcsolatok feltérképezése. Ehhez az egyes tantárgyakra vonatkozó leíró statisztikai módszerek mellett elengedhetetlen, hogy olyan eszközöket is alkalmazzunk, amelyek ezen kapcsolatok feltárására alkalmasak: a korrelációt és a regressziót.

4.1.4.1. Korreláció

A hallgatók különböző tantárgyakból elért eredményei közötti kapcsolat vizsgálatára a korrelációs együttható alkalmazható. Korrelációs vizsgálatok célja az, hogy megmutassuk két ismerv közötti kapcsolat szorosságát és irányát. Az érdemjegyek, illetve a teljesítmény-metrika sorrendi skálán történő értelmezése révén a Pearson-féle korrelációs együttható nem alkalmas a kapcsolat mérésére, hiszen az egyes érdemjegyek bár számértékek, valójában csak sorbarendezi tudjuk őket. Helyette a sorrendi skálát, az adatok egymáshoz viszonyított sorrendjét figyelembe vevő Spearman-féle rangkorrelációs együttható használható. Ennek lényege, hogy az ismervértékekhez sorszámokat rendel, majd az így kapott rangsorok közötti kapcsolatot vizsgálja az alábbi képlet szerint, ahol R_X és R_Y jelöli a két ismerv szerinti sorszámokat, N pedig a darabszámot.

$$\rho = 1 - \frac{6 \cdot \sum (R_X - R_Y)^2}{N \cdot (N^2 - 1)}$$

Amennyiben a rangkorrelációs együttható 0 értéket vesz fel, nincs kapcsolat a két ismerv között, 1 esetén pozitív, -1 esetén pedig negatív függvénytípusú kapcsolat van (Zar, 2005) (Hauke–Kossowski, 2011).

4.1.4.2. Regresszió

Ahhoz, hogy meg tudjuk állapítani, hogy a teljesítmény-metrikák értékeit milyen egyéb változók és milyen mértékben magyarázzák, többváltozós regressziós elemzéseket kell elvégeznünk. Mivel az új metrikánk, azaz a regressziós modell eredményváltozója sorrendi skálán értelmezett változó, ezért nem használható a klasszikus lineáris regresszió, hiszen ez arra ad becslést, hogy az eredményváltozó várható értéke a magyarázó változók hatására hogyan változik. Mivel a várható érték sorrendi skálán nem értelmezhető, ezért ebben az esetben egy speciálisabb regressziós módszert, a Least Absolute Deviation regressziót (LAD regresszió) kell alkalmazni. Ez a klasszikus megközelítéssel ellentétben nem a feltételes átlagot becsli, hanem a feltételes medián értékeket és nem a négyzetes hibaösszeget, hanem az abszolút hibaösszeget minimalizálja az együtthatók megválasztásakor. Ezzel egységnyi eltérést kisebb mértékben büntet, mint a négyzetes hibafüggvény, figyelembe véve, hogy sorrendi skálán egységnyi eltérésnek kevésbé szabad jelentőséget tulajdonítani. Mediánt sorrendi skálán is van értelme számítani megfelelő szélességű értékkészlet mellett, ami a teljesítmény-metrikának köszönhetően a rendelkezésünkre áll (Hunyadi–Vita, 2006) (Fitzenberger–Koenker–Machado, 2013).

A megfelelő regressziós modell kialakításához – tehát amiben csak az eredményváltozóra szignifikáns hatást gyakorló magyarázó változók vannak – backward változószelekciós eljárást érdemes alkalmazni. Ennek lényege, hogy kezdetben minden ismert változót magyarázó változóként beveszünk a regressziós modellbe, majd ezekre vonatkozó parciális próbákkal, és ezek p-értékeinek segítségével meghatározzuk és eltávolítjuk a legkevésbé szignifikáns magyarázó változót a modellből. A parciális próbák nullhipotézise azt állítja, hogy az adott magyarázó változó regressziós együtthatója 0, vagyis nincs magyarázó ereje a változónak a sokaságban. Így ennek szerepeltetése a regresszióban felesleges és esetleg káros is a modell általánosító képességére nézve. Alternatív hipotézise pedig, hogy az adott magyarázó változó regressziós együtthatója szignifikáns mértékben különbözik a 0-tól a sokaságban, így a változó releváns a modellben. A legkevésbé szigni-

fikáns magyarázó változó eltávolítása után újra elkészítjük a regressziós modellt. Ezeket a lépéseket addig ismételjük, amíg már csak szignifikáns magyarázó változók maradnak a modellben (Hunyadi–Vita, 2008).

A p-érték alapú szelekció azonban feltételezi a multikollinearitás hiányát (azaz a magyarázó változók függetlenségét), illetve, hogy a hibatag homoszkedasztikus. Ezek nem teljesülése esetén a p-értékek felfelé torzulnak, így, ha kizárólag ezek alapján szelektálunk, lényeges változókat hagyhatunk ki a modellből. Ezért nem elegendő csak a p-érték alapján döntenünk arról, hogy meddig ismételjük újra és újra az eljárást, hanem meg kell figyelni a modell jóságát mutató Akaike információs kritérium (Akaike information criterion, AIC) értékét is. Ez a mutató csökken, ha az összesített modellhiba (klasszikus regresszió esetén a négyzetes hibaösszeg, LAD regressziónál pedig az összesített abszolút hiba) csökken, de nő, ha új változó kerül bevonásra a modellbe jelentős modellhiba csökkenés nélkül. Tehát ez egy olyan kritérium, ami a modell illeszkedését a magyarázó változók számát büntetve nézi, ezzel megelőzve a modell túlillesztését. Amennyiben ez az érték egy magyarázó változó eltávolításával emelkedni kezd, a modell futtatását le kell állítani (Venables–Ripley, 2002) (Hunyadi–Vita, 2006).

Annak a vizsgálatára, hogy egy tárgyból elért eredményre hatással van-e minőségi ismerv – a mi esetünkben ez a nem és a hallgató évfolyama –, a változókat a regressziós modellbe dummy változóként kell felvenni. Ehhez a változók kimenetei közül választunk egy referenciaértéket, majd azt vizsgáljuk, hogy ehhez képest mennyivel tér el az eredményváltozó értéke egy másik kimenetnél. Kettőnél több lehetséges kimenet esetén, ha a változószelekciós eljárás során nem találjuk szignifikánsnak a dummy változó egyikét, akkor azt a kimenetet eltávolítva a modellből ugyanúgy tovább futtathatjuk az algoritmust. Ez úgy értelmezhető, hogy a dummy változó referenciaértékét és a nem szignifikáns kimenetet összevonjuk (Hunyadi–Vita, 2006).

5. OSZTÁLYZATELEMZÉS A GAZDASÁGINFORMATIKA KÉPZÉSBEN

Dolgozatunk célja olyan osztályzatelemzési keretrendszer kialakítása, amely a felsőoktatási képzések struktúrájának fejlesztését segíti. A megfelelő eszközök kiválasztásához egy konkrét kutatás kapcsán haladtunk végig a tanulóelemzés folyamatán. Ennek során arra törekedtünk, hogy feltárjuk az osztályzatelemzés lehetőségeit a Budapesti Corvinus Egyetem Gazdaságinformatikus BSc képzésén, olyan megoldásokat keresve, amelyekkel az oktatás minőségét javítani lehet. Így kezdetben feltáró jelleggel fogtunk hozzá az elemzési munkálatoknak, azzal a céllal, hogy kapcsolatokat találjunk a hallgatók egyes tantárgyakban nyújtott teljesítményei között, illetve minél több információt tudjunk kinyerni a Neptun rendszerben eltárolt, hallgatói eredményekre vonatkozó adatokból. Jól látható tehát, hogy egy nem strukturált problémával álltunk szemben, melynél sem jól meghatározott cél, sem az ahhoz vezető pontos út nem volt ismert. Ezért a vizsgálatok során nem egy konkrét tanulóelemzési modellt alkalmaztunk, hanem a 3. fejezetben ismertetett általános folyamat elemein haladtunk végig.

5.1. A RENDELKEZÉSRE ÁLLÓ ADATOK BEMUTATÁSA

A kiinduló adatok forrásrendszere a Budapesti Corvinus Egyetem Neptun rendszere volt. Rendelkezésünkre álltak három táblázatban a tanulmányaikat 2011/2012, 2012/2013, illetve 2013/2014 őszi félévében kezdett gazdaságinformatikus hallgatók azonosítói. Ezeket közvetlenül a Neptun rendszerből történő adatkinyerést követően kaptuk meg, a hallgatók beazonosításra alkalmas Neptun kódok hash-kóddá alakításával. Emellett megkaptuk a 2011-es év őszi félévétől kezdve tizenhárom matematikai és informatikai tárgyból – Analízis, Informatika alapjai, Valószínűségi számítás, Számítástudomány, Lineáris algebra, Statisztika I., Statisztika II., Szoftver-technológia I., Szoftver-technológia II., Operációkutatás, Adatbázis rendszerek, Informatikai rendszerek fejlesztése, Üzleti intelligencia – elért hallgatói eredményeket attól a félévtől kezdve, ahogyan azt a 2011/2012-ben kezdett évfolyam legelőször tanulhatta, egészen addig a félévig, amíg a 2013/2014-es évfolyam hallgatói utoljára részt vehettek azokon. Ettől eltérő eset, hogy Üzleti intelligenciából a legkésőbbi rendelkezésre álló adat a 2018/2019 őszi féléves volt. Ebből kifolyólag a 2013/14-ben kezdett évfolyamban lehetnek olyanok, akiknek a Budapesti Corvinus Egyetem Tanulmányi és Vizsgaszabályzat (továbbiakban: TVSZ) III./50§ 2. és III./70§ 5. pontok szerint megengedett maximális 4 tárgyfelvevételig még lett volna két próbálkozási lehetőségük. A nyers adattáblákban rendelkezésre álló adatok a hallgatók neve és évfolyama, érdemjegyre vonatkozó utolsó bejegyzés

és annak típusa (vizsgajegy, gyakorlati jegy, aláírás stb.) voltak, illetve, ha volt ilyen, akkor a korábbi eredményük az adott tárgyból, annak típusa és az ahhoz tartozó dátum (lásd 1. táblázat).

1. táblázat: A tantárgyakhoz tartozó táblázatok kezdeti állapota

Kód	Nem	Bejegyzés értéke	Bejegyzés típusa	Korábbi eredmény
e05d9ae7317f9cf644c0ed78927b35cd06c97650	Férfi	Elégséges	Vizsgajegy	Elégséges(Vizsgajegy-2012.01.24)
882481d03ca0440868e261ba975e5976b7ee843a	Férfi	Közepes	Vizsgajegy	Közepes(Vizsgajegy-2012.01.10)
de644e6e75fe5df07a393c4d9319bf9fc1687901	Férfi	Elégséges	Vizsgajegy	Elégséges(Vizsgajegy-2012.01.27),Elégtelen(Vizsgajegy-2012.01.24)
6392b580d7f36cedce489b3158c3f37ae2fb46a3	Férfi	Közepes	Vizsgajegy	Közepes(Vizsgajegy-2012.01.03)
82c726cd370bcf9d4598c6baf7513f775c6b7115	Férfi	Közepes	Vizsgajegy	Közepes(Vizsgajegy-2011.12.23)
7b91f6ac760cf8bb556828b7934712be5744cb29	Férfi	Közepes	Vizsgajegy	Közepes(Vizsgajegy-2012.01.17)
f5e86c6066c6a5d71c6ea6c5d0b6032306ddc10b	Nő	Jó	Vizsgajegy	Jó(Vizsgajegy-2011.12.23)

Forrás: saját szerkesztés

6. AZ ADATOK ELŐKÉSZÍTÉSE

Az adatok összegyűjtését követően az elemzéseket lehetővé tevő formára kellett hoznunk ezeket. Az adattisztítási és -átalakítási folyamat során a legfontosabb szem előtt tartott tényező az volt, hogy egységes, jól kezelhető, minden szükséges kinyerhető információt tartalmazó adattáblákat kapjunk. A lépésekhez használt eszköz a Microsoft Excel volt.

6.1. A KAPOTT ATTRIBÚTUMOK

Az adatelőkészítés során tantárgyanként vizsgáltuk, hogy a hallgatók milyen bejegyzéseket kaptak a rendelkezésre álló forrás adattáblákban (lásd 1. táblázat). A kihívást az jelentette, hogy az egy évfolyamba tartozó hallgatók nem feltétlenül egy félévben vették fel a tárgyat. Emellett azoknak az eredményeit, akik felvették a kurzust, de mégsem vizsgáztak, illetve azoknak, akik elégtelen érdemjegyet szereztek, több év adataiból kellett kiszűrni és összerendezni egy táblázatba. Ezeket figyelembe véve összefésültük és átalakítottuk a tantárgyi és a hallgatói adatokat. Ennek eredményeként mindhárom évfolyamra egy-egy olyan adattábla állt elő tantárgyanként, amelyben szerepelt a hallgatók egyedi azonosítója, neve, a végső eredménye sorrendi skálán értelmezett érdemjeggyé kódolva, illetve a jegy típusa, azaz, hogy vizsgajegyet, gyakorlati jegyet vagy aláírást kaptak az adott tárgyból. Továbbá abból, hogy az adott tárgyat tekintve hány félév adataiban szerepelt egy hallgatói azonosító, kiszámoltuk a hozzá tartozó tárgyfelvételek számát, illetve a bejegyzések számából elvéve a „Nem vizsgázott” és a „Megtagadva” bejegyzéseket megkaptuk a vizsgafelvételek számát. A vizsgafelvételek számából kivéve a „Nem jelent meg” és az „Igazoltan távol” bejegyzések számát, megkaptuk, hogy hányszor vizsgázott a hallgató, a tárgy elvégzésének éve pedig az

utolsó bejegyzéshez tartozó év lett. Azoknak a hallgatóknak, akiknek végső eredménye „Nem vizsgázott”, „Nem jelent meg”, „Megtagadva”, illetve „Igazoltan távol” volt, végső jegyüket 0-ként határoztuk meg, hiszen ők a próbálkozások ellenére semmilyen eredményt nem értek el az adott tárgyból. Gyakorlati jegyes kurzusok esetében az egységes forma¹ miatt vizsgázások számaként jelöltük azt, hogy a hallgató hányszor kapott valamilyen bejegyzést, függetlenül attól, hogy ilyen esetben nem vizsga keretei között, hanem szorgalmi időszakban született meg az eredmény.

Az egységes formátumra hozás eredményeképpen állt elő a 2. táblázatban látható adatstruktúra évfolyamonként minden tantárgyra.

2. táblázat: A tisztított adattáblák felépítése

Hallgatói azonosító	Nem	Végső bejegyzés	Végső jegy	Bejegyzés típusa	Tárgyfelvételek száma	Vizsgafelvételek száma	Vizsgázások száma	Végzés éve	
ef58b610ee835da7bcf8d15e33b54b3cda032d14	1	Elégséges		2	Vizsgajegy	1	1	1	11-12
c896cfaabbc8a937076c4ef536bd5b375d2f2b9f	0	Jeles		5	Vizsgajegy	1	1	1	11-12
a92664d1aa708641fe9ffc2886f3a7823116757a	0	Jó		4	Vizsgajegy	1	1	1	11-12
b0713ca0b9d1f8c964ab3cdb07e89391160e5fc6	1	Jó		4	Vizsgajegy	1	2	2	11-12
de20a106996519080399c82c8765502f61a415a0	1	Jó		4	Vizsgajegy	1	1	1	11-12
5ab9700c9ead46d32319982e724fe3b8db081df7	1	Elégséges		2	Vizsgajegy	1	2	2	11-12
bcbbbd75973caf5b9aed8d1ee76aec731926e73e	1	Közepes		3	Vizsgajegy	2	2	2	13-14
086e6865844296654dda486043ebb205d26e561b	1	Nem jelent meg		0	Vizsgajegy	2	2	0	13-14
↓	↓	↓	↓	↓	↓	↓	↓	↓	
nominális	nominális	ordinális	ordinális	nominális	arány	arány	arány	ordinális	

Forrás: saját szerkesztés

A megfelelő elemzési módszerek kiválasztásához fontos volt meghatároznunk, hogy az így kapott attribútumokat milyen skálán értelmezzük. Ahogy az a 2. táblázatban is látható, a hallgatói azonosító, a nem és a végső bejegyzés típusa nominális skálán mért adatok, a végső jegy és a végzés éve ordinális, a tárgyfelvételek, vizsgafelvételek, illetve vizsgázások száma pedig arány skálán mért attribútumok.

6.2. ANOMÁLIÁK KEZELÉSE

Az adatainkat a Neptun tanulásirányítási rendszerből nyertük ki, ami jelentősen megszabja az adatbevitel módját. Habár a jegyek tekintetében minimális az adatminőségi hibák lehetősége, az itt tárolt adatok többsége jelentős mennyiségű anomáliát tartalmaz. Kutatásunk ezen pontján ezeket a problémákat igyekeztünk feltárni, és megoldást kidolgozni a kezelésükre.

¹ Az egységes forma alatt azt értjük, hogy gyakorlati jegyes tárgynál a hallgató végső bejegyzése (ami lehetett érdemjegy, „Nem vizsgázott”, vagy „Megtagadva” bejegyzés) vizsgázásként legyen rögzítve, hogy az elemzésnél ne legyen eltérés a gyakorlati és vizsgajegyes tárgyaknál és ugyanaz alatt az oszlopelnevezés alatt szerepelhessen mindkettő esetében.

Az első és legszembetűnőbb probléma a hiányzó adatokkal kapcsolatban merült fel. Kiinduláskor rendelkezésünkre álltak évfolyamonként a hallgatók listái, illetve tantárgyanként az elért eredmények. Azonban mikor a hallgatói névsorhoz rendeltük a kurzusokhoz kapcsolódó változókat, minden esetben előfordult olyan hallgató, akihez az adott tárgy eredményei között nem találtunk adatot. Ez azt jelentette, hogy ezek a személyek fel sem vették a kurzust. A hiányzó adatok két típusát azonosítottunk be. Az első eset a könnyebben kezelhető: olyan hallgatók, akikhez egyik tárgy kapcsán sem találtunk bejegyzést. Ők azok, akiket bár felvettek, és bekerültek egy adott évfolyamba, végül nem kezdték el az egyetemet. Mivel a gazdaságinformatika képzésben résztvevők eredményeit szeretnénk volna elemezni, így egyértelmű volt, hogy ezeket a hallgatókat kiszűrjük az adatok közül. A 2011/2012-es évfolyamban 4, a 2012/2013-asban 3, míg a 2013/2014-es évfolyamban 5 ilyen tanuló volt. A másik eset, amikor egy hallgató bizonyos tárgyakból szerzett bejegyzést, azonban volt legalább egy olyan kurzus, amit nem vett fel, így ehhez kapcsolódóan nem voltak adatai. Ezeket az eseteket ebben a fázisban még nem szűrtük ki, azonban a későbbiekben ez még további problémákat jelentett. Ezek kezelésére és megoldásukra a 6.3.4-es pontban térünk vissza.

Ezenfelül több esetben is előfordult, hogy egy-egy hallgató azonosítója egynél több évfolyamban jelent meg. Ez abból adódhatott, hogy az adott tanuló újra felvételizett, és így két vagy akár több évfolyamba is bekerült a neve. Számunkra ez azért jelentett gondot, mert az adattisztítás végső célja egy aggregált, elemzéshez jól használható adatbázis létrehozása volt. Ehhez azonban elengedhetetlen, hogy minden adat csak egyszer szerepeljen, vagyis a duplikációkat ki kellett szűrni, hogy egy-egy hallgatóról pontosan beazonosítható legyen, hogy melyik tárgyból milyen érdemjegyet szerzett. Megoldásként mérlegeltük az újrafelvételiző hallgatók adatainak összevonását és egy sorban letárolását, azonban ez további hibalehetőségeket vetett volna fel. Így előfordulhatott volna, hogy a tárgyfelvételek számánál 4-nél nagyobb érték szerepel, ami már értékhatár hibát eredményezett volna a TVSZ szerint. Ezért felmértük, hogy hányszor fordult elő ez a helyzet, és azt kaptuk eredményül, hogy évfolyamonként csak 2-2 ilyen eset volt. Emiatt úgy döntöttünk, hogy ezeket az értékeket kivesszük az adatbázisból, és csak azokkal a hallgatókkal dolgozunk tovább, akik csak egyszer szerepelnek a táblákban.

Miután kiszűrtük ezeket a problémás eseteket, további anomáliákat fedeztünk fel az adatok kapcsán. A gazdaságinformatikus szakhoz tartozó operatív tanterv megadja, hogy melyik tárgyat melyik szemeszterben érdemes felvenni, és ezek hogyan épülnek egymásra. Előfordultak azonban

olyan esetek, amikor egy hallgató a tantervtől eltérő módon egy adott kurzust olyan félévben teljesített, amikor az még nem lett volna számára lehetséges. Mivel ekkorra már kiszűrtük az újrafelvételizőket, így ezt az okot kizártuk. A probléma forrásának pontos beazonosítása ebben az esetben nehezebb volt. Az egyik lehetséges magyarázat az, ha a hallgató másik szakon már teljesítette az adott tárgyat, és az átjelentkezésnél elfogadták az érdemjegyet. Így a bejegyzés rögzítésének ideje a korábbi teljesítés időpontja lett. Ebben az esetben a későbbi elemzéseknél nem az jelentett volna problémát, hogy az adatokat nem tudjuk használni a számításokhoz, hanem az értelmezések kapcsán adhatott volna nem valós képet. Hiszen, ha a gazdaságinformatikus képzésről szeretnénk elmondani valamit, akkor az eredményt torzítna az, ha a hallgatók másik szakon szerzett érdemjegyet is bele vesszük a számításokba. Ezért amellet döntöttünk, hogy ezeket a hallgatói eredményeket is kiszűrjük az adatbázisból. Az eltávolított értékek darabszámát és arányát az összes adathoz képest (ez már nem tartalmazza a hiányzó adatokat és az újrafelvételizőket) tárgyanként összesítve a három évfolyamra a 3. táblázat tartalmazza:

3. táblázat: Operatív tanterven kívüli tárgyfelvelelek száma és aránya az összes adathoz képest

Tantárgy	Operatív tanterven kívüli tárgyfelvelelek száma	Arány
Analízis	0	0%
Informatika	0	0%
Statisztika I.	1	0,19%
Lineáris algebra	0	0%
Számítástudomány	0	0%
Szoftver-technológia I.	0	0%
Valószínűségyszámítás	2	0,38%
Adatbázis rendszerek	4	0,75%
Operációkutatás	5	0,94%
Statisztika II.	5	0,94%
Szoftver-technológia II.	1	0,19%
Informatikai rendszerek fejlesztése	2	0,38%
Üzleti intelligencia	3	0,56%

Forrás: saját szerkesztés

Összességében az adatminőségi hibák közül a hiányzó adatok kezelése, a duplikációk kiszűrése és a konzisztencia megtartása jelentette a legnagyobb kihívást. A fent ismertetett három eset kezelésével az volt a célunk, hogy pontosan definiálni tudjuk, hogy egy sor mit jelent az adattáb-

lánkban. Ehhez tudnunk kell, hogy az adott hallgató melyik évfolyamba tartozik, és abban is biztosnak kell lennünk, hogy a tárgyakat a gazdaságinformatika képzés keretein belül végezte el. Jelen esetben ennek biztosítása azt jelentette, hogy az anomáliákat tartalmazó sorokat kivettük az adathalmazunkból, ezt pedig azért tudtuk megtenni, mert a problémás esetek aránya minimális volt.

6.3. ADATSTRUKTÚRA KIALAKÍTÁSA – AGGREGÁLÁS

Miután az adatminőségi hibákat feltártuk, és a problémákat megoldottuk, a következő lépésben egy aggregált, további elemzésekre alkalmas adatstruktúra kialakításán dolgoztunk. Évfolyamonként és tantárgyanként külön adattáblák álltak már a rendelkezésünkre a 4. ábrán bemutatott mezőkkel. A célunk tehát az volt, hogy egy adatbázisba kerüljön az összes adat úgy, hogy ne vesszen el információ, de duplikációk se merüljenek fel.

Az aggregálást alapvetően két szinten kellett megoldanunk: az egyes hallgatókhoz hozzárendelni az összes tantárgyhoz tartozó értékeiket és a tantárgyak mentén összekapcsolni az évfolyamokat.

6.3.1. Aggregálás a hallgatók mentén

Ebben a lépésben azt szeretnénk volna elérni, hogy egy sorban legyenek letárolva az egyes hallgatóknak mind a tizenhárom tárgyra vonatkozó adatai. Egyértelmű volt, hogy a Neptun kódból képzett hash-kód lesz az az érték, azonosító, aminek segítségével össze tudjuk fűzni a különálló tábláinkat (ez egyedi azonosító, tehát egyértelműen meghatározza az egyedet). Azonban az attribútumok neveit át kellett alakítanunk úgy, hogy a későbbiekben meg tudjuk különböztetni az egyes kurzusokra vonatkozó értékeket. Ennek megfelelően a mezőneveket kiegészítettük egy-egy előtaggal, amelyeket a tárgynevek rövidítéseiből kaptunk (1. számú melléklet). Ezen lépés kivitelezéséhez az Excel keretein belül elérhető Power Query alkalmazást használtuk.

6.3.2. Aggregálás az évfolyamok mentén

A fenti lépés elvégzése után már csak három adattáblánk volt, ami a három évfolyam adatait tartalmazta. Ezek összefűzésénél az összekapcsolás alapját már nem a hallgatók jelentették, hiszen ezek az értékek évfolyamonként eltérőek, hanem az egyes tantárgyak. Tehát míg az előző lépésben oszlopokat kapcsoltunk össze, most a sorokat kellett egymás alá fűznünk. Ehhez szintén a Power Query alkalmazást használtuk, illetve arra kellett még figyelniünk, hogy létrehozzunk egy plusz mezőt, ami az évfolyam változót tartalmazza.

6.3.3. A végleges adatstruktúra

Az adataggregálás során előállt egy olyan adatbázis, ami alapvetően három dimenzióra épül: hallgatói azonosító (sorok azonosítója), tantárgyak (oszlopok azonosítója) és az évfolyam (mező). Ennek megfelelően bármelyik cellát kiválasztva egyértelműen meghatározható, hogy az adott érték melyik hallgatóhoz, melyik tantárgyhoz és melyik évfolyamhoz tartozik. A 4. táblázat mutatja be, hogy hogyan néz ki az adatbázis, és hogyan jelennek meg a különböző dimenziók.

4. táblázat: Példa a végleges adatstruktúrára

Hallgatói azonosító	Nem	Évfolyam	ui_szovjegy	ui_típus	ui_targyfelv	ui_vizsgafelv	ui_vizsga	ui_ev	ui_jegy
ef58b610ee835da7bcf8d15e33b54b3cda032d14	Férfi	11-12	Jó	Gyakorlati jegy	1	1	1	2014-2015	4
f7770d133fb120a84fcabfaf1b34c150ca23b50e	Nő	11-12	Közepes	Gyakorlati jegy	1	1	1	2014-2015	3
f20a01387537c1b61766f73ac6ddaf1a71d05ed6	Férfi	11-12	Jó	Gyakorlati jegy	1	1	1	2014-2015	4
11d268f0d2ca283f2f2b02ac4d3afa5dbd3250aa	Nő	11-12	Jó	Gyakorlati jegy	1	1	1	2014-2015	4
40489226825faa81e88d06dfd92c654859e5f13c	Férfi	11-12	Jeles	Gyakorlati jegy	1	1	1	2014-2015	5
e52780a87efe6006f1efd6037caec4fcb4ada4b8	Férfi	11-12	Jó	Gyakorlati jegy	1	1	1	2014-2015	4

Forrás: saját szerkesztés

6.3.4. Hiányzó adatok kezelése az aggregált adatbázisban

A 6.3.1. pontban ismertetett Neptun kód szerinti összefűzés során azonban további nehézségekkel találkoztunk. Több esetben is előfordult, hogy bizonyos tárgyakat egyes hallgatók fel sem vettek, így hozzájuk semmilyen érték nem tartozott az adott táblában. A JOIN feltételnél így meg kellett határoznunk, hogy ez INNER vagy OUTER típusú legyen. A mi esetünkben ez azt jelentette, hogy csak azokat a hallgatókat akarjuk-e megtartani, akiknek minden tantárgyhoz volt valamilyen bejegyzésük (INNER), vagy azokat is bele vesszük, akiknél ez feltétel nem teljesül (OUTER). Ez utóbbi eset további kérdéseket is felvetett, hiszen ekkor a hiányzó mezőkre egységesen meg kell határozni, hogy üresen hagyjuk, vagyis NULL értékűek lesznek, vagy megpróbálunk egy becsült, közelítőleges értéket adni. Az 5. táblázatban látható, hogy tárgyanként, évfolyam szerinti bontásban mekkora volt a hiányzó értékek aránya.

5. táblázat: Hiányzó adatok aránya

Tantárgy	11-12	12-13	13-14
Analízis	14,41%	9,31%	10,43%
Informatika	8,62%	4,50%	4,39%
Statisztika I.	10,34%	6,53%	6,58%
Lineáris algebra	13,56%	7,84%	8,26%
Számítástudomány	8,47%	7,35%	8,70%
Szoftver-technológia I.	16,95%	12,75%	15,22%
Valószínűségszámítás	20,69%	12,12%	15,79%
Adatbázis rendszerek	18,10%	13,07%	15,56%
Operációkutatás	20,34%	13,24%	16,96%
Statisztika II.	22,41%	14,14%	17,78%
Szoftver-technológia II.	21,19%	16,18%	17,83%
Informatikai rendszerek fejlesztése	22,41%	17,50%	21,24%
Üzleti intelligencia	24,58%	16,53%	24,78%

Forrás: saját szerkesztés

Vizsgálatunk során arra jutottunk, hogy három esetet tudunk elkülöníteni, de nincs igazán jó megoldás, mivel mindegyiknek vannak negatív hatásai. Éppen ezért a kutatásunk ezen fázisában egyiket sem vetettük el, hanem a későbbiekben mindhárommal dolgoztunk tovább. Így az elemzések eredményei mellett azt is vizsgáltuk, hogy végső soron melyik adatbázis lenne a legmegfelelőbb.

Az első, amire a későbbiekben *Szűrt* adatbázisként fogunk hivatkozni, csak azokat az egyedekeket tartalmazza, amelyek esetében a hallgatóknak az összes tárgyhöz kapcsolódóan van adatuk. Ezzel azonban nagyon sok sort kiszelektálunk, egészen pontosan az összes adat 27,10%-át. Ez a későbbiekben torzíthatja az elemzések eredményeit, hiszen feltételezzük, hogy az esetek nagy részében olyan hallgatókról van szó, akik elégtelen érdemjegyet szereztek bizonyos kurzusokból, és ezért nem tudtak más tárgyakat felvenni. Ezzel a megközelítéssel vélhetően pont a leggyengébb hallgatók adatait dobjuk el. A későbbiekben ezt a felvetést is megvizsgáltuk, és számadatokkal igyekeztünk alátámasztani.

Másik megközelítésként az összes adatot megtartottuk olyan formában, ahogyan azok az anomáliák kiszűrése után a rendelkezésünkre álltak, vagyis az üres mezők NULL értéket vettek fel. Ebben az esetben egyáltalán nem veszik el adat, azonban sok bizonytalanságot okoz az, hogy a különböző számítási módszerek hogyan reagálnak a NULL adattípusra. A későbbiekben erre *Összes null* adatbázisként hivatkozunk.

Harmadik lehetőségként szintén megtartottuk az összes adatot, azonban a NULL adattípust lecseréltük, hogy biztosan ne okozzon gondot a későbbiekben. Az *Összes* néven hivatkozott adatbázisban az érdemjegyek esetében 0-s értéket adtunk az összes olyan hallgatónak, aki nem vette fel a tárgyat. Ez azonban szintén adhat hamis vagy torzított eredményeket, mert csak egy feltételezést tartalmaz arra vonatkozóan, hogy akik nem vették fel a tárgyat, azok nem is teljesítenék.

7. A KIDOLGOZOTT OSZTÁLYZATELEMZÉSI MEGOLDÁSOK BEMUTATÁSA

Az adatok előkészítése után a következő lépésben ezek elemzésével foglalkoztunk. Első körben az adatok mélyebb megismerése volt a cél, így leíró statisztikai módszerekkel igyekeztünk átfogó képet szerezni az érintett tizenhárom tárggyal a kapcsolatban. Ezután olyan új metrika kidolgozását tűztük ki célul, amely lehetővé teszi, hogy egy ismérvben kezeljük a hallgatók egyes tantárgyakban nyújtott teljesítményét leíró három ismérvet (jegy, tárgyfelvételek és vizsgázások száma). Ezt felhasználva pedig olyan elemzési módszerek halmazát állítottuk össze, mely nem csak a tárgyak közötti kapcsolatokat, de az egyéb változók összefüggéseit is feltárja.

Az egyes elemzési lépések végén az Informatikai rendszerek fejlesztése és az Üzleti intelligencia tárgyak segítségével mutatjuk be, hogy a kapott eredményeket a gyakorlatban hogyan tudjuk értelmezni. Azért ezeket a kurzusokat választottuk, mert ez a kettő a gazdaságinformatika képzés utolsó féléveiben vehető fel, és az alapozó tárgyak során elsajátított tudásra építenek.

7.1. LEÍRÓ STATISZTIKA

A leíró statisztikai elemzések során az *Összes null* adatbázissal dolgoztunk, hiszen a célunk az volt, hogy megismerjük az összes rendelkezésre álló adatot. A *Szűrt* adatbázis ebből a szempontból nem volt használható, az *Összes* pedig becsült adatokat is tartalmazott, melyek torzítottak volna a leíró jellegű eredményeken. Az elemzéseket a Microsoft Excel, az R-3.5.3. (R Core Team, 2019) és az RStudio (RStudio Team, 2020) segítségével végeztük el, melyhez a *readexcel* (Wickham–Bryan, 2019) és a *dplyr* (Wickham–François–Henry–Müller, 2020) csomagokat használtuk fel. A vizualizációk elkészítéséhez pedig a Tableau szoftvert alkalmaztuk.

7.1.1. Hiányzó és szelektált adatok

Az elemzések során első körben azokat a problémás adatokat vizsgáltuk, melyeket kiszűrtünk az adatelőkészítéskor. A 6. táblázat mutatja, hogy évfolyamonként hány főt vettek fel a gazdaságinformatikus képzésre éves bontásban. Ebből kiszűrve az újrafelvételizőket és azokat, akik el sem

kezdték a képzést, megkaptuk a szűrt elemszámot. Habár csak három évfolyamot vizsgáltunk, ezekről elmondható, hogy növekvő tendenciát mutat az egy évfolyamra felvett hallgatók száma. 2013-ban 99,1%-kal többen kezdték meg tanulmányaikat a képzésen, mint két évvel korábban.

6. táblázat: Rendelkezésre álló és kiszűrt adatok száma

	Elemszám	Újrafelvételiző	Nem kezdte el	Szűrt elemszám
11-12	118	2	4	112
12-13	204	4	3	197
13-14	230	2	5	223

Forrás: saját szerkesztés

Az így kapott elemszámokat tovább vizsgáltuk tantárgyi bontásban. A 7. táblázat kurzusonként mutatja meg, hogy hány hiányzó adat volt, illetve a *Problémás* oszlopban azoknak a hallgatóknak a száma szerepel, akik olyan félévben teljesítették a tárgyat, amikor ez nem lett volna lehetséges. Ezeket kisselektálva kaptuk meg a szűrt elemszámot, vagyis azokat a hallgatókat, akikkel a leíró statisztikai elemzések során ténylegesen foglalkozni tudtunk.

7. táblázat: Elemszám tantárgyi bontásban

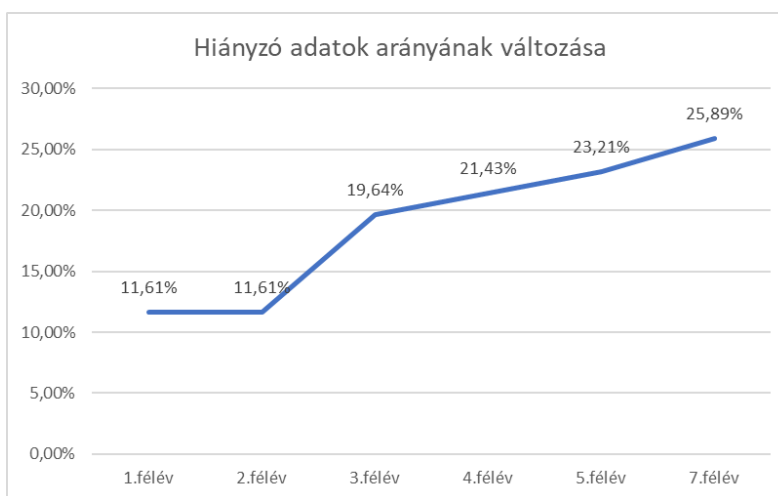
Tantárgy	11-12				12-13				13-14			
	Elemszám	Problémás	Hiányzó adat	Szűrt elemszám	Elemszám	Problémás	Hiányzó adat	Szűrt elemszám	Elemszám	Problémás	Hiányzó adat	Szűrt elemszám
Analízis	112	0	17	95	197	0	19	178	223	0	24	199
Informatika	112	0	10	102	197	0	9	188	223	0	10	213
Statisztika I.	112	0	12	100	197	1	13	183	223	0	15	208
Lineáris algebra	112	0	16	96	197	0	16	181	223	0	19	204
Számítástudomány	112	0	10	102	197	0	15	182	223	0	20	203
Szoftver-technológia I.	112	0	20	92	197	0	26	171	223	0	35	188
Valószínűségszámítás	112	0	24	88	197	2	24	171	223	0	36	187
Adatbázis rendszerek	112	0	21	91	197	1	26	170	223	3	35	185
Operációkutatás	112	0	24	88	197	2	27	168	223	3	39	181
Statisztika II.	112	0	26	86	197	2	28	167	223	3	40	180
Szoftver-technológia II.	112	0	25	87	197	0	33	164	223	1	41	181
Informatikai rendszerek fejlesztése	112	0	26	86	197	0	35	162	223	2	48	173
Üzleti intelligencia	112	0	29	83	197	1	20	176	223	2	54	167

Forrás: saját szerkesztés

Szembetűnő volt, hogy a félévek sorrendjébe rendezett tantárgyakon lefelé haladva egyre nő a hiányzó adatok száma, ezért ezt további módszerekkel is megvizsgáltuk. Rendelkezésünkre állt az operatív tanterv, amely rögzíti, hogy melyik kurzust melyik szemeszterben ajánlott felvenni. Az itt bemutatott táblázatok is ezt a sorrendet követik, és az egyes félévekhez tartozó tárgyak háttér-színezése azonos. Megvizsgáltuk tehát, hogy az egyes szemeszterekbe tartozó kurzusoknál mekkora volt a hiányzó adatok arányának átlaga, majd ezeket az intenzitási viszonyszámokat a 2. ábrán

jelenítettük meg (a vizsgált kurzusok között nem állt rendelkezésünkre 6. féléves tárgy, ezért hiányzik ez az érték). A vonaldiagram egyértelműen megmutatja a növekvő tendenciát a vizsgált változók között a 2. félévtől kezdve, vagyis szemeszterről szemeszterre morzsolódnak le hallgatók, és az utolsó féléves tárgyat (Üzleti intelligencia) már csak a hallgatók 74,11%-a kezdte el.

2. ábra: Hiányzó adatok arányának változása

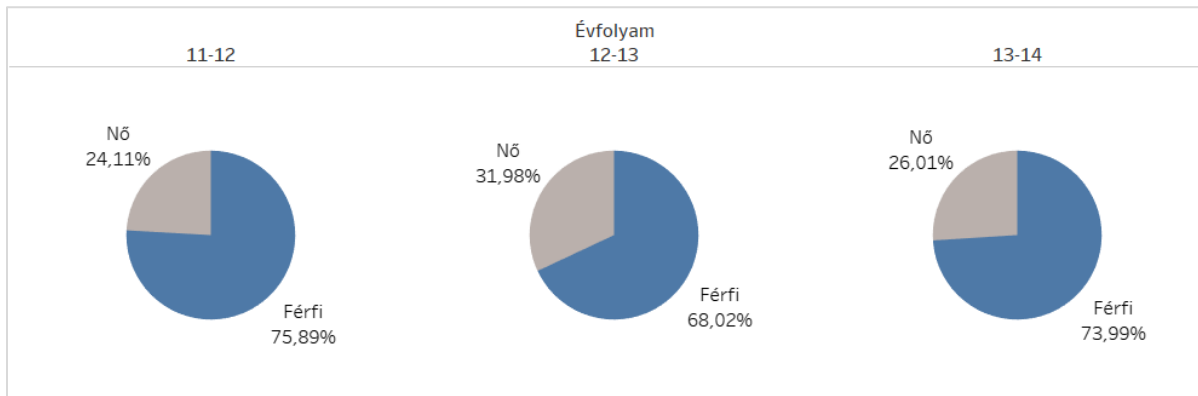


Forrás: saját szerkesztés

7.1.2. A változók vizsgálata

Ezt követően a nemek arányát vizsgáltuk a képzésen belül évfolyamokra bontva. Ahogy azt a 3. ábrán a kördiagramok szemléltetik, a 2011/2012-es és 2013/2014-es évfolyamokban a nők aránya közelítőleg 25% volt, míg a 2012/2013-ban ez az érték 7%-ponttal magasabb. Azonban így is az mondható el a gazdaságinformatikus képzésről, hogy ebben a három évfolyamban a férfiak létszáma nagyjából háromszorosa a nők létszámának.

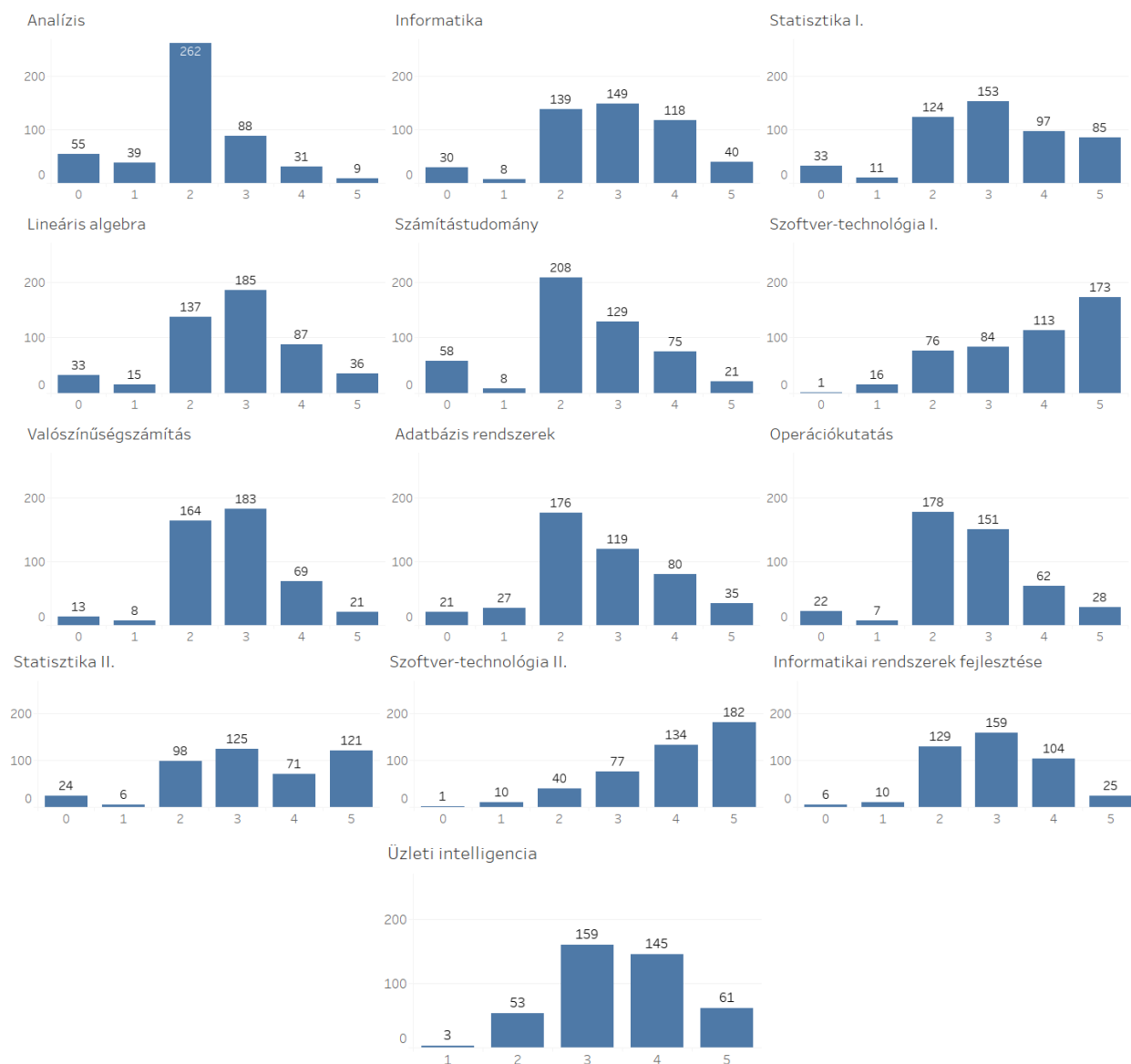
3. ábra: Nem szerinti megoszlás



Forrás: saját szerkesztés

A végső bejegyzés, vagyis az elért érdemjegy kapcsán az értékek eloszlását vizsgáltuk, azonban ebben az esetben a 3 évfolyamot összesítve elemeztük. A 4. ábrán láthatunk példát jobbra és balra elnyúló eloszlásokra is, azonban pár esetben közel normális eloszlásra is találtunk példát. Többek között az Analízis, az Adatbázis rendszerek, az Operációkutatás és a Számítástudomány tárgyak számítanak az adatok alapján a nehezebben teljesíthető kurzusok közé, ezekben az esetekben kevés hallgatónak sikerült jobb érdemjegyet szereznie. Ezzel szemben a Szoftver-technológia I. és II. esetében a hallgatók nagyobb számban szereztek jeles bejegyzést, a bukások aránya jelentősen kisebb. Az Informatika, az Informatikai rendszerek fejlesztése és az Üzleti intelligencia kurzusoknál megközelítőleg normális eloszlás figyelhető meg.

4. ábra: Érdemjegyek eloszlása



Forrás: saját szerkesztés

A tárgyfelvételek számának vizsgálata során felfedeztünk olyan eredményeket (lásd 8. táblázat), melyek összhangban vannak az érdemjegyek során megállapított felvételekkel. Habár minden kurzus esetén a legtöbben csak egyszer jelentkeztek az adott tárgyra, az átlag és szórás tekintetében vannak jelentősebb különbségek. Két szélsőséges eset, az Analízis és Üzleti intelligencia mutatóin mutatjuk be az eredményeket. Az Üzleti intelligencia tárgyfelvételeinek száma esetében az átlag 1,01, amitől a hallgatók átlagosan minimálisan tértek el: a szórás értéke 0,1. Ezenfelül a ferdeség és csúcsosság mutatók nagyon magas értékűek, vagyis egy szélsőségesen jobbra elnyúló

csúcsos eloszlásról van szó. Ezzel szemben az Analízis esetén az átlag már magasabb érték, vagyis a hallgatók átlagosan 1,49-szer vették fel a tárgyat, illetve a felvételek száma is jobban szórt. A ferdeség és csúcsosság mutatók még ebben az esetben is pozitívak, vagyis továbbra is egy csúcsos, jobbra elnyúló eloszlásról beszélünk. Azonban mivel értékük kisebb, ezért ebben az esetben már nincsenek nagymértékben kiugró adatok és darabszámok, az eloszlás egyenletesebb. A két tárgy közötti jelentős eltérés oka az lehet, hogy az Üzleti intelligencia utolsó féléves tárgy, így az a hallgató, aki eljutott idáig, már ezt az akadályt is feltételezhetően könnyebben át tudta lépni, ellenben az 1. féléves Analízissel.

8. táblázat: Tárgyfelvételek számának vizsgálata

	Tárgyfelvétel						
	Minimum	Maximum	Módusz	Átlag	Szórás	Ferdeség	Csúcsosság
Analízis	1	5	1	1,49	0,81	1,53	1,48
Informatika	1	4	1	1,06	0,32	6,17	42,02
Statisztika I.	1	3	1	1,08	0,34	4,31	19,01
Lineáris algebra	1	3	1	1,16	0,42	2,61	6,45
Számítástudomány	1	4	1	1,33	0,66	1,88	2,47
Szoftver-technológia I.	1	4	1	1,12	0,43	3,99	-0,80
Valószínűségyszámítás	1	4	1	1,21	0,55	2,66	6,48
Adatbázis rendszerek	1	4	1	1,43	0,71	1,51	1,38
Operációkutatás	1	5	1	1,22	0,57	2,86	8,79
Statisztika II.	1	5	1	1,22	0,58	2,95	9,46
Szoftver-technológia II.	1	3	1	1,07	0,26	3,91	15,33
Informatikai rendszerek fejlesztése	1	3	1	1,13	0,39	3,22	10,28
Üzleti intelligencia	1	2	1	1,01	0,10	10,11	101,47

Forrás: saját szerkesztés

A vizsgázások számának vizsgálata során kapott eredményeket a 9. táblázat foglalja össze. Ezen értékek értelmezése hasonló a korábbi adatokéhoz, jelen esetben is az Analízis és az Üzleti intelligencia jelentik a két szélsőséget. Az összes tárgyról elmondható, hogy a legtöbb hallgató mindössze egy alkalommal vett fel vizsgát, azonban az átlag, szórás, ferdeség és csúcsosság mutatók között vannak jelentősebb eltérések. Az eredmények alapján szintén azt a következtetést lehet levonni, hogy a matematikai tárgyakat nehezebben teljesítették a hallgatók, több vizsgaalkalomra volt szükségük. Azonban itt fontos kiemelni, hogy a kurzusokat két kategóriába soroljuk: vizsgajegy és gyakorlati jegy tárgyak. Utóbbi esetén a vizsgázások száma alacsonyabb, hiszen ezeket a hallgatók évközben teljesítik, nincs szükségük vizsgaalkalmakra. Ezt szemlélteti a 5. ábra, ahol egyértelműen látszik, hogy a vizsgajegy tárgyak esetében a vizsgázások számának átlaga magasabb. Hasonló következtetést lehet levonni a többi mutató kapcsán is, hiszen a gyakorlati jegy

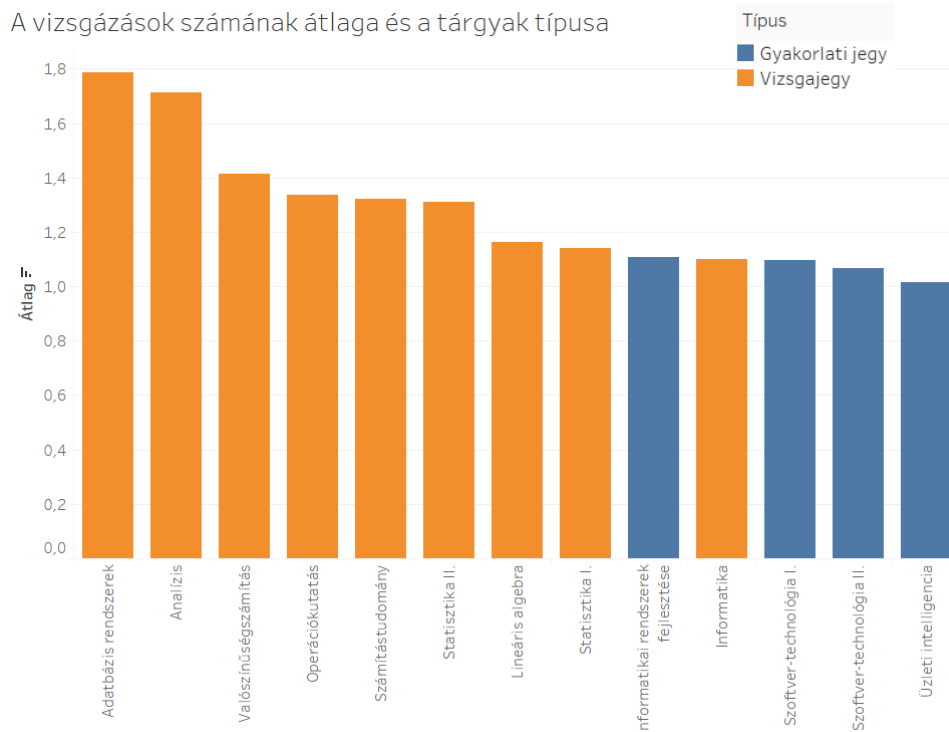
tárgyak esetén a szórás kisebb, és csúcsosabb, nagyobb mértékben jobbra elnyúló eloszlásokat figyelhetünk meg.

9. táblázat: A vizsgázások számának vizsgálata

	Vizsgázások						
	Minimum	Maximum	Módusz	Átlag	Szórás	Ferdeség	Csúcsosság
Analízis	0	7	1	1,71	1,32	1,50	2,39
Informatika	0	7	1	1,10	0,52	3,66	34,45
Statisztika I.	0	5	1	1,14	0,60	2,31	11,00
Lineáris algebra	0	5	1	1,16	0,59	1,82	6,53
Számítástudomány	0	6	1	1,32	0,98	2,01	5,00
Szoftver-technológia I.	0	4	1	1,10	0,38	4,14	-0,80
Valószínűességszámítás	0	6	1	1,41	0,94	2,52	6,68
Adatbázis rendszerek	0	9	1	1,79	1,17	1,68	4,41
Operációkutatás	0	6	1	1,34	0,75	1,90	6,32
Statisztika II.	0	5	1	1,31	0,72	1,83	4,70
Szoftver-technológia II.	0	3	1	1,07	0,26	3,50	14,00
Informatikai rendszerek fejlesztése	0	3	1	1,11	0,39	2,73	9,89
Üzleti intelligencia	1	2	1	1,02	0,13	7,56	55,83

Forrás: saját szerkesztés

5. ábra: A vizsgázások számának átlaga és a tárgyak típusa



Forrás: saját szerkesztés

7.1.3. A leíró statisztikai eredmények ismertetése az IRF és ÜI kurzusokon keresztül

A leíró statisztikai elemzések eredményeit összesítve az Informatikai rendszerek fejlesztése és az Üzleti intelligencia tantárgyakról elmondható, hogy a képzésre felvett hallgatók közelítőleg 75%-a kezdte el ezeket a kurzusokat. Azonban azon hallgatók számára, akik eljutottak idáig, többnyire nem okozott gondot ezeknek a tárgyaknak a teljesítése, hiszen az IRF esetében a tanulók 3,8%-ának nem sikerült legalább elégséges érdemjegyet szereznie, míg az ÜI esetében ez az arány mindössze 0,7%. Ez valóban jól reprezentálja, hogy akik a korábbi féléves alapozó tárgyakat sikeresen teljesíteni tudják, azoknak ezek a kurzusok sem okoztak problémát. Esetleg felmerülhet a gondolat, hogy ennek a valódi oka az, hogy ezek a tárgyak könnyebbek a többihez képest. Ezt a felvetést azonban cáfolja a 4. ábrán bemutatott érdemjegyek eloszlása. Jól látszik, hogy mind a két esetben közel normális eloszlásról beszélhetünk, ami azt jelenti, hogy viszonylag kevés hallgatónak sikerült kiváló érdemjegyet szereznie, a legtöbben közepes bejegyzéssel zárták a kurzust. Mivel gyakorlati jegyes tárgyakról van szó, így a tárgyfelvételek és a vizsgázások száma között nincs jelentős eltérés, amit a 8. és 9. táblázat mutatói is bizonyítanak. Az Üzleti intelligencia tárgyat a 3 évfolyamban mindössze 4 hallgatónak kellett kétszer felvenni, a többieknek sikerült elsőre teljesíteni. Az Informatikai rendszerek fejlesztése esetében ez az érték magasabb, 37-en kétszer, 9-en háromszor próbálkoztak. Összességében azonban elmondható, hogy a hallgatók számára ezek nemcsak egy-egy könnyen teljesíthető tárgyat jelentenek, hanem kihívást is. A két kurzus valóban képes különbséget tenni a hallgatók között, és csak kevesen tudják jeles érdemjeggyel zárni, de ugyanígy kevesen vannak azok is, akik egyáltalán nem tudják teljesíteni. Ez pedig nagyon fontos hiszen az utolsó félévekben a hallgatók számára már kritikus, hogy ne kelljen egy-egy kurzust újból felvenniük, hiszen ez akár plusz féléveket is jelenthet számukra az egyetemen.

A leíró statisztikai elemzések végére megismertük a változóink eloszlását, azt, hogy milyen értékkészleten mozognak, és mik számítanak tipikus és kiugró értékeknek. Ennek eredményeképp nemcsak a rendelkezésre álló adatokról kaptunk egy átfogó képet, hanem arról is, hogy az adott három évfolyamban mely tárgyak számítottak nehezebben teljesíthetőnek a jegyek, a tárgyfelvételek és a vizsgázások számának tekintetében.

7.2. METRIKA KIALAKÍTÁSA

7.2.1. A metrika szükségessége

Az érdemjegyek szűk értékkészlete miatt, ahogyan azt az előző pontban is láthattuk, a statisztikai elemzések végrehajtása sok esetben kizárólag szemléltetésre alkalmazható (hisztogram), a később tervezett elemzésekhez pedig kifejezetten nem használható. Ezért fontosnak tartottuk, hogy kialakítsunk egy olyan új teljesítmény-metrikát, amely szélesebb értékkészleten tudja elhelyezni a hallgatói teljesítményt, mindezt úgy, hogy az érdemjegyen kívül más változókkal is bővíti, mélyíti annak jelentését. Megvizsgáltuk, hogy a tisztított, rendelkezésre álló adatok közül mely attribútumok azok, amelyek relevánsak és jól beépíthetők ebbe a metrikába. Az érdemjegyek mellett mindenképpen fontos információt hordoz egy tantárgy elvégzésével kapcsolatban, hogy a hallgató hány tárgyfelvétellel és hány vizsgázással érte el az adott eredményt. Gyakorlati jegyes tárgynál a kiugró esetektől eltekintve – amikor a hallgató a vizsgaidőszakban javítási vagy pótlási lehetőséget kapott – egy tárgyfelvételhez egy vizsgázás és egy vizsgafelvétel tartozhat. Így, ahogyan ezt már a 7.1.2. pontban említettük, több próbálkozással teljesített tárgyak esetén már eleve különbség lesz gyakorlati jegyes és vizsgajegyes tantárgyak között, mivel az utóbbihoz kevesebb vizsgázás tartozhat. Ezt az eltérést még jobban növelné, ha nemcsak a vizsgázások, hanem a vizsgafelvételek számát is számításba vennénk a metrika kialakításakor, ezért ezt a változót a továbbiakban nem vizsgáltuk.

7.2.2. Elméleti kérdések a kialakítandó metrikával kapcsolatban

A metrika kialakítása előtt végig kellett gondolnunk annak felépítésével kapcsolatos kérdéseket. Elsőként meg kellett határoznunk, hogy jutalmazó vagy büntető modellt alkalmazzunk. Azaz, az érdemjegyek értékét a tárgyfelvétel- és vizsgaszámok növekedésével csökkentjük, vagy a lehető legtöbb próbálkozással elért érdemjegyek legyenek az adott jegyhez tartozó minimális értékek és minden ennél kevesebb próbálkozást jutalmazunk. Mivel a vizsgált adathalmazban voltak olyan hallgatók, akiknek a TVSZ-ben meghatározott maximális tárgyfelvétel- és vizsgaszámnál több próbálkozásuk volt, ezzel kiszámíthatatlanná téve ezen attribútumok értékkészletét, ezért úgy döntöttünk, hogy a büntető módszert alkalmazzuk, hiszen ellenkező esetben nem lehetne biztosra mondani, hogy mi az a maximális próbálkozásszám, amihez a minimális metrika értéket rendeljük érdemjegyenként. Természetesen erre egy lehetséges megoldás, ha a TVSZ-ben megengedett maximális próbálkozások számánál nagyobbakat értékhatár hibaként kiszűrjük az adatok tisztításánál, de ezzel pont azoknak a hallgatóknak az eredményeit távolítanánk el az adathalmazból, akiknek

különösen nagy nehézséget okozott a tárgy teljesítése, ezzel mesterségesen javítanánk az adott tantárgy összeredményét. Ezzel szemben a büntető modell esetében az 1 vizsgázás és 1 tárgyfelvétel kombinációhoz azok a metrika értékek tartoznak, amelyek a ténylegesen elért érdemjegyeknek felelnek meg. Így ebben az esetben egyértelműen meghatározható a maximális metrika érték, az 5,0.

A próbálkozások számának büntetése miatt jogosan vetődik fel a kérdés, hogy megszabhatunk-e egy olyan értéket – jelen esetben a 0-t – aminél biztosan nem lehet kisebb a metrika értéke. Ehhez azt kellett végiggondolni, hogy lehetséges-e annyi próbálkozással elérni elégtelen érdemjegyet, vagy nem is kapni végső eredményt a tárgyra, aminél már azt mondhatjuk, hogy ekvivalens azzal az állapottal, amikor a hallgató fel sem vette a tárgyat. Ahogy azt majd a későbbiekben láthatjuk, erre a kérdésre nem tudtunk egyértelműen válaszolni, így a kialakított háromféle metrikából az egyiket úgy hoztuk létre, hogy negatív értéket is felvehessen rossz érdemjegy és próbálkozás-szám kombináció esetében.

Ezután arra a kérdésre kerestük a választ, hogy vajon egyenlő mértékben kell-e büntetni az egyes érdemjegyek esetében az újabb próbálkozások számát. E kérdés mögött az a pedagógiai megközelítés áll, hogy aki sokadik próbálkozásra jól teljesíti a tárgyat, az a végső érdemjegyéhez képest kevesebbet érdemel, mint aki sokadik próbálkozásra sem teljesíti, vagy kevésbé jó végeredménnyel zárja a kurzust. Például aki felvette háromszor a tárgyat és ötször elment vizsgázni, majd jeles érdemjegyet ért el, annak a jegye ugyanolyan szintű ötösnek számít-e, mint amilyen szintű elégségesnek mondható annak az eredménye, aki ugyanennyi próbálkozással kap elégséges jegyet. Jogosnak tartjuk azt a felvetést, hogy jobb érdemjegynél nagyobb mértékben büntessük a próbálkozások számát, mint egy kevésbé kiemelkedőnél, ezért a végső modellek közül kettőt e szerint a logika szerint alakítottunk ki.

Ehhez szorosan kapcsolódó kérdés, hogy egy sokadik próbálkozásra megszerzett jó jegy lehet-e rosszabb, mint egy kevesebb próbálkozással szerzett kevésbé jó eredmény. Arra jutottunk, hogy például egy négy tárgyfelvétellel és hat vizsgával elért jeles érdemjegy kevesebbet ér, mint egy első próbálkozásra megszerzett jó bejegyzés. Annak a finomhangolására viszont, hogy ez milyen mértékben és meddig igaz, csak a metrika konkrét kialakításánál térünk majd ki.

Gyakorlati jegyes kurzusoknál minden tárgyfelvevételhez maximum egy vizsga tartozhat (az említett szélsőséges esetektől eltekintve). Így a vizsgajegyes tárgyakhoz képest a hallgatói eredmények összességében jobbak lennének, mesterségesen jobb eredményeket generálnánk. Ezért úgy döntöttünk, hogy gyakorlati jegyes tárgyak esetében a vizsgázások számát jobban büntetjük.

7.2.3. A metrika kialakításához alkalmazott eszközök

A fenti kérdéseket figyelembe véve háromféle metrikát alakítottunk ki. Mindegyik alapja olyan exponenciális modell, melynek logikája, hogy az érdemjegyek értékét a vizsgázások és tárgyfelvevételek súlyozott összegével bünteti. A modellek általános képlete az alábbiakban látható, ahol x jelöli az elért érdemjegyet, y a tárgyfelvevételek számát, a ennek a kitevőjét, z a vizsgázások számát, b pedig ennek a kitevőjét:

$$x + 1 - \frac{(1,5y^a + 0,5z^b)}{2}$$

Már itt is megfigyelhető, hogy a kitevőkön kívül az y 1,5-es és a z 0,5-es szorzójával is különbséget teremtettünk a próbálkozás típusa között – jobban büntetve a tárgyfelvevételt.

A következő szakaszban a modellek kialakításához alkalmazott eszközök bemutatása következik.

7.2.3.1. Érzékenységvizsgálat

A modellek finomhangolásához, és a kitevők alakulásának vizsgálatára érzékenységvizsgálatot alkalmaztunk, amihez az Excel what-if-analysis módszerét használtuk. Az összes tárgyfelvevétel és vizsgaszám kombinációra figyeltük, hogy a különböző nagyságú kitevők esetében az egyes érdemjegyekhez milyen értékeket rendelnek a metrikák. A 10. táblázatban látható szürke sávban határoztuk meg azokat az értékeket, amikhez tartozó kitevő-kombinációt elfogadhatónak gondoltunk. Ehhez elsősorban azt a gondolatot tartottuk szem előtt, hogy a vizsgázások számát kisebb mértékben büntetjük, mint a tárgyfelvevételek számát, mivel újabb tárgyfelvevételkor a hallgatónak újra kell hallgatnia a tárgyat. Illetve arra is törekedtünk, hogy a kitevők között ne legyen nagymértékű eltérés, hiszen akkor aránytalanul nagyobb mértékben büntetnénk a tárgyfelvevételek számát. Emellett a táblázatban sötétebb szürkével látszik az a terület, amelyben található értékeket ideálisnak véltünk, sárgával pedig a végleges döntésünket jelöltük.

10. táblázat: Érzékenységvizsgálat a metrika értékeire közepes érdemjegy esetén

Tárgyfelvétel	Vizsga											
2	2											
		Vizsga kitévő										
	1,875	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00	1,10	1,20
Tárgyfelvétel kitévő	0,20	2,85	2,83	2,81	2,78	2,76	2,73	2,70	2,67	2,64	2,60	2,56
	0,30	2,79	2,77	2,75	2,72	2,70	2,67	2,64	2,61	2,58	2,54	2,50
	0,40	2,72	2,70	2,68	2,66	2,63	2,60	2,58	2,54	2,51	2,47	2,44
	0,50	2,65	2,63	2,61	2,59	2,56	2,53	2,50	2,47	2,44	2,40	2,36
	0,60	2,58	2,56	2,53	2,51	2,48	2,46	2,43	2,40	2,36	2,33	2,29
	0,70	2,49	2,47	2,45	2,43	2,40	2,38	2,35	2,32	2,28	2,25	2,21
	0,80	2,41	2,39	2,36	2,34	2,32	2,29	2,26	2,23	2,19	2,16	2,12
	0,90	2,31	2,29	2,27	2,25	2,22	2,19	2,17	2,13	2,10	2,06	2,03
	1,00	2,21	2,19	2,17	2,15	2,12	2,09	2,06	2,03	2,00	1,96	1,93
	1,10	2,11	2,08	2,06	2,04	2,01	1,99	1,96	1,93	1,89	1,86	1,82
	1,20	1,99	1,97	1,95	1,92	1,90	1,87	1,84	1,81	1,78	1,74	1,70

Forrás: saját szerkesztés

7.2.3.2. Táblázatok az értékkészlet vizsgálatára

A metrika értékeinek monitorozására a különböző modellek esetében a 11. táblázathoz hasonló táblázatokat alakítottunk ki. Ezek minden érdemjegy esetén az összes lehetséges jegy, tárgyfelvétel- és vizsgaszám kombinációra tartalmazzák a hozzájuk tartozó modell értékeket. Habár a TVSZ négy tárgyfelvételben és hat vizsgázásban határozza meg a maximális próbálkozások számát, látva a vizsgált adathalmazban előforduló, ezektől eltérő eseteket, megjelentítettünk a megengedett értéknél nagyobbakat is. Ezek a táblázatok segítettek a modell finomhangolásánál megfigyelni a különböző értékkombinációk esetében a metrika értékkészletének alakulását. Segítségével tudtuk többek között a 7.2.2. alfejezetben kifejtett kérdéses eseteket figyelni: hogy a metrikák értéke lehet-e 0 alá, vagy hogy egy sokadik próbálkozásra szerzett jobb érdemjegy lehet-e rosszabb egy kevesebb próbálkozással elért rosszabb érdemjegynél, illetve, hogy abban az esetben, amikor nem egyenletes mértékben büntetünk, akkor mekkora lehet az eltérés az érdemjegyek között.

11. táblázat: Értékkészlet vizsgálata jó érdemjegyre vonatkozóan

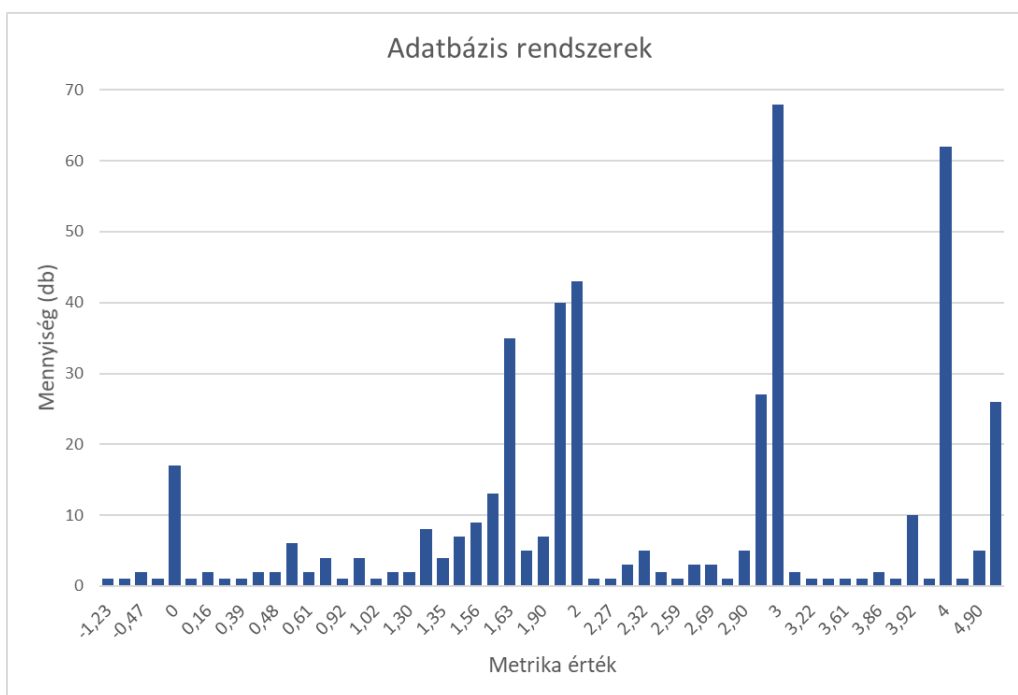
Tárgyfelvétel	Vizsga	Jegy	a) modell	b) modell	c) modell
1	1	4	4,000	4,000	4,000
2	1	4	3,689	3,613	3,613
3	1	4	3,451	3,300	3,300
4	1	4	3,250	3,027	3,027
5	1	4	3,073	2,780	2,780
1	2	4	3,942	3,881	3,920
2	2	4	3,632	3,495	3,533
3	2	4	3,393	3,182	3,220
4	2	4	3,192	2,908	2,947
5	2	4	3,015	2,662	2,700
1	3	4	3,902	3,787	3,862
2	3	4	3,592	3,401	3,475
3	3	4	3,353	3,088	3,162
4	3	4	3,152	2,814	2,889
5	3	4	2,975	2,568	2,642
1	4	4	3,871	3,707	3,815
2	4	4	3,560	3,320	3,428
3	4	4	3,322	3,007	3,115
4	4	4	3,121	2,734	2,842
5	4	4	2,944	2,487	2,595
2	5	4	3,534	3,248	3,387
3	5	4	3,296	2,934	3,074
4	5	4	3,095	2,661	2,801
5	5	4	2,918	2,414	2,554
2	6	4	3,511	3,181	3,351
3	6	4	3,273	2,868	3,038
4	6	4	3,072	2,595	2,765

Forrás: saját szerkesztés

7.2.3.3. Oszlopdiagramok és gyakorisági táblák az értékkészlet és az eloszlások vizsgálatára

Szintén az értékkészlet, illetve az egyes tárgyak esetében a metrika értékek gyakorisági eloszlásának vizsgálatára gyakorisági táblákat és oszlopdiagramokat készítettünk, hogy a finomhangolás során minél jobban figyelemmel kísérhessük, hogy egy-egy változás milyen mértékű eltérést okoz a vizsgált adathalmazon. Ahogyan az a 6. ábrán található oszlopdiagramon is látszik, és a többi vizsgált tantárgy esetében is megfigyelhető, az értékek móduszai főként az egész érdemjegyeknél találhatók, azaz legtöbbször egy tárgyfelvétellel és egy vizsgázással teljesítették a tárgyakat, amint azt már a 7.1.2. pontban is említettük.

6. ábra: Metrika értékek gyakorisági eloszlása oszlopdiagramon



Forrás: saját szerkesztés

7.2.4. Gyakorlati jegyes tárgyak

Ahogy arra a 6.1. pontban már kitértünk, a gyakorlati jegyes tárgyaknál szintén vizsgázásnak kódoltuk az érdemjegy rögzítését, függetlenül attól, hogy a vizsgajegyes tárgyakkal ellentétben gyakorlati jegy esetén a hallgató egy félévben csak egyszer próbálkozhatott, azaz egy tárgyfelvevélhez maximum egy vizsgázás tartozott (kivéve a szélsőséges eseteket). Érthető tehát, hogy mivel a vizsgázások számát a modellek a metrika értékek csökkentésével büntetik, ezért az így kapott vizsgajegyes tárgyakból elért eredmények jelentősen rosszabbak lennének, mint a gyakorlati jegyesek. Ezért a modelleket gyakorlati jegyes tárgyak esetében is elő kellett állítani, úgy, hogy az első vizsgát követően minden újabb vizsgázást nagyobb mértékben büntessünk, mint azt a vizsgajegyes tárgyaknál tennénk.

7.2.5. A három metrika

A 7.2.2. pontban ismertetett kérdésekkel és a 7.2.3. pontban bemutatott eszközök használatával kialakítottunk három olyan modellt, amelyek a teljesítmény-metrikák alapját jelentik, és megfelelnek az összetett hallgatói teljesítmény mérésére. Az így kialakított modelleket még egy változóval finomítani kellett. Az alábbiakban ezek részletes ismertetése következik.

7.2.5.1. Az a) modell

Ezen modell segítségével kialakított metrika minden érdemjegy esetében ugyanolyan mértékben bünteti a próbálkozások számát. Így törekedtünk arra is, hogy a TVSZ-ben maximálisan megengedett próbálkozásszámot meghaladó esetektől eltekintve a metrika szerint kapott érték ne lehessen rosszabb, mint az eredetileg kapott érdemjegy. Azaz például egy közepes jegyhez tartozó metrika-érték ne lehessen 2,0 vagy az alatti. Azonban a gyakorlati jegyes tárgyaknál a nagyobb mértékű büntetés miatt a metrika értéke szélsőséges esetben kicsivel a valós érdemjegyek alá csökkenhet. Ez azonban csak olyankor fordulhat elő, ha a tárgyfelvételek száma 4, vizsgázások száma pedig 6. Gyakorlati jegyes tárgyaknál azonban az esetek nagy részében a tárgyfelvételek számánál a vizsgázások száma nem nagyobb, kivéve azokat a ritka helyzeteket, amikor a hallgató a vizsgaidőszakban javítási lehetőséget kapott. Ez azonban a vizsgált adatok között, egy tárgyon belül, ugyanannál a hallgatónál egynél többször nem fordult elő. Ebből következik az is, hogy a 0-s érdemjegyként kódolt eredményektől eltekintve – amelyek a „Nem vizsgázott”, „Nem jelent meg”, „Megtagadva” bejegyzések – negatív értéket nem vehet fel az ilyen módon kialakított függvény. Vizsgajegyes tárgyaknál az általános képlet az alábbiak szerint módosul:

$$x + 1 - \frac{(1,5y^{0,5} + 0,5z^{0,3})}{2}$$

Ebben a modellben $a=0,5$ és $b=0,3$, ezzel az y és z szorzóin túl a kitevőkkel is súlyoztuk a próbálkozásokat, még nagyobb mértékben függővé téve a büntetés mértékét a próbálkozás típusától (tárgyfelvétel vagy vizsga). A modell sajátossága, hogy minden érdemjegy esetében ugyanakkora a büntetés mértéke. A megengedett maximális próbálkozásszámot nem meghaladó teljesítménynél a vizsgajegyes tárgyaknál legfeljebb 0,928-del lehet kisebb a metrika értéke az eredeti érdemjegynél. Gyakorlati jegyes tárgyaknál is hasonló gondolatmenet mentén alakítottuk a metrikát, azzal az eltéréssel, hogy ebben az esetben a vizsgázások számát nagyobb mértékben büntetjük, $b=0,5$ -del. Így itt a hagyományos esetekben maximum 1,112-del kevesebb a metrika érték, mint az érdemjegy (de ahogy már említettük, ilyen eset a vizsgált adatok között nem fordult elő).

$$x + 1 - \frac{(1,5y^{0,5} + 0,5z^{0,5})}{2}$$

Az alábbi táblázatban összefoglaljuk a gyakorlati jegyes és a vizsgajegyes értékek eltérésének mértékét:

**12. táblázat: Az a) metrika értékkészlete gyakorlati és vizsgajegy
tárgyak esetében**

Tárgyfelvétel	Vizsga	Jegy	Vizsgajegy	Gyakorlati jegy
1	1	3	3,000	3,000
2	1	3	2,689	2,689
3	1	3	2,451	2,451
4	1	3	2,250	2,250
5	1	3	2,073	2,073
1	2	3	2,942	2,896
2	2	3	2,632	2,586
3	2	3	2,393	2,347
4	2	3	2,192	2,146
5	2	3	2,015	1,969
1	3	3	2,902	2,817
2	3	3	2,592	2,506
3	3	3	2,353	2,268
4	3	3	2,152	2,067
5	3	3	1,975	1,890
1	4	3	2,871	2,750
2	4	3	2,560	2,439
3	4	3	2,322	2,201
4	4	3	2,121	2,000
5	4	3	1,944	1,823
2	5	3	2,534	2,380
3	5	3	2,296	2,142
4	5	3	2,095	1,941
5	5	3	1,918	1,764
2	6	3	2,511	2,327
3	6	3	2,273	2,089
4	6	3	2,072	1,888

Forrás: saját szerkesztés

7.2.5.2. A b) modell

A második modell kialakításánál a célunk az volt, hogy szem előtt tartsuk azt a véleményt, miszerint egy sokadik próbálkozásra elért jeles érdemjegy már nem feltétlen mondható jelesnek, viszont egy hasonló módon szerzett elégséges még továbbra is elégségesnek felel meg. Azaz jobb érdemjegynél nagyobb mértékben büntettük a próbálkozások számát úgy, hogy a kitevőket súlyoztuk a kapott érdemjegy értékével is. Így minél jobb érdemjegyet kapott a hallgató, annál jobban büntettük az ehhez felhasznált próbálkozásai számát. Ezen kívül vizsgajegy tárgyak esetén az általános képlet úgy módosult, hogy $a=0,15x$ és $b=0,14x$ lett:

$$x + 1 - \frac{(1,5y^{0,15x} + 0,5z^{0,14x})}{2}$$

Gyakorlati jegyes tárgyaknál pedig az alábbiak szerint kerültek meghatározásra a metrika értékek. Bár ebben az esetben $a=0,15x$ és $b=0,2x$, x és y szorzójának köszönhetően ez a modell is nagyobb mértékben bünteti a tárgyfelvételek számát.

$$x + 1 - \frac{(1,5y^{0,15x} + 0,5z^{0,2x})}{2}$$

A kétféle tárgytípusra kapható értékek eseteinek összevetése közepes és jeles érdemjegyek esetében a 13. táblázatban látható:

13. táblázat: A b) metrika értékeinek alakulása gyakorlati és vizsgajegyes tárgyak esetében

Tárgyfelvétel	Vizsga	Jegy	Vizsgajegy	Gyakorlati jegy	Jegy	Vizsgajegy	Gyakorlati jegy
1	1	3	3,000	3,000	5	5,000	5,000
2	1	3	2,725	2,725	5	4,489	4,489
3	1	3	2,520	2,520	5	4,040	4,040
4	1	3	2,350	2,350	5	3,629	3,629
5	1	3	2,203	2,203	5	3,242	3,242
1	2	3	2,916	2,871	5	4,844	4,750
2	2	3	2,641	2,597	5	4,333	4,239
3	2	3	2,436	2,391	5	3,884	3,790
4	2	3	2,266	2,222	5	3,473	3,379
5	2	3	2,118	2,074	5	3,086	2,992
1	3	3	2,853	2,767	5	4,711	4,500
2	3	3	2,579	2,492	5	4,199	3,989
3	3	3	2,374	2,287	5	3,751	3,540
4	3	3	2,204	2,117	5	3,339	3,129
5	3	3	2,056	1,969	5	2,953	2,742
1	4	3	2,802	2,676	5	4,590	4,250
2	4	3	2,528	2,401	5	4,079	3,739
3	4	3	2,323	2,196	5	3,631	3,290
4	4	3	2,153	2,026	5	3,219	2,879
5	4	3	2,005	1,878	5	2,832	2,492
2	5	3	2,484	2,319	5	3,967	3,489
3	5	3	2,279	2,114	5	3,519	3,040
4	5	3	2,109	1,944	5	3,107	2,629
5	5	3	1,961	1,796	5	2,721	2,242
2	6	3	2,445	2,243	5	3,862	3,239
3	6	3	2,240	2,038	5	3,414	2,790
4	6	3	2,070	1,868	5	3,002	2,379

Forrás: saját szerkesztés

Megfigyelhető, hogy míg egy közepes érdemjegy esetében a maximálisan megengedett próbálkozásokat meg nem haladók esetében nagyjából egy érdemjegynyi romlott az érték, addig jeles érdemjegy esetében a sokadik (de a megengedettnél nem több alkalommal történő) próbálkozásra megkapott jegy körülbelül egy első próbálkozásra megszerzett közepes érdemjegynek felel meg.

7.2.5.3. A c) modell

Ezen modell segítségével meghatározott metrika értékek kialakítása a másik kettőtől különböző módon történik. A modellben a és b értéke a $b)$ modellhez hasonlóan nem minden jegy esetében egyezik meg, ám értékük nem a kapott érdemjegy és valamilyen érték szorzata, hanem egy külön táblázatban lettek kialakítva az egyes érdemjegyekhez tartozó súlyok, így könnyebben hangolhatóvá téve a modellt.

14. táblázat: Kitevőket tartalmazó táblázat a c) modell esetében

Érdemjegy	Tárgyfelvétel kitevő	Vizsgázás kitevő
0	0,7	0,5
1	0,6	0,4
2	0,5	0,3
3	0,5	0,3
4	0,6	0,4
5	0,7	0,5

Forrás: saját szerkesztés

A külső táblázatban történő szerkesztés lehetővé tette azt, hogy úgy büntessük eltérő mértékben az egyes érdemjegynél felhasznált próbálkozások számát, hogy a büntetés mértéke nem csökken folyamatosan a kapott érdemjegyekkel. Ezzel tudtuk azt megoldani, hogy kiemelkedőbb eredményeknél és az elégtelen, vagy 0-s jegy esetén is jobban büntetjük a plusz próbálkozást. Azaz, ha egy hallgató sokadik próbálkozásra sem kapott érdemjegyet, azt is jobban megbüntettük, mint a közepes eredményt elérő hallgatókat, hiszen ő gyakorlatilag feleslegesen használt kapacitást. Vizsgajegyes tárgyak esetében a metrika értékét meghatározó modellhez az általános képletet alkalmaztuk.

$$x + 1 - \frac{(1,5y^c + 0,5z^d)}{2}$$

Azonban ebben az esetben c -vel a 14. táblázatban az adott érdemjegyhez tartozó tárgyfelvétel kitevőt, d -vel pedig a vizsgázás kitevőt jelöltük. A megfelelő súly megtalálásához az Excel FKERES függvényét használtuk. Gyakorlati jegyes tárgyak esetében d egy 1,3-szeres szorzót kap, ezzel növelve a vizsgázások súlyát:

$$x + 1 - \frac{(1,5y^c + 0,5z^{1,3d})}{2}$$

Az alábbi táblázatban láthatók a c) metrika értékei a különböző tárgyfelvétel, vizsga és érdemjegy kombinációkra:

15. táblázat: A c) metrika értékkészlete elégtelen és közepes érdemjegyre

Tárgyfelvétel	Vizsga	Jegy	Vizsgajegy	Gyakorlati jegy	Jegy	Vizsgajegy	Gyakorlati jegy
1	1	1	1,000	1,000	3	3,000	3,000
2	1	1	0,613	0,613	3	2,689	2,689
3	1	1	0,300	0,300	3	2,451	2,451
4	1	1	0,027	0,027	3	2,250	2,250
5	1	1	-0,220	-0,220	3	2,073	2,073
1	2	1	0,920	0,892	3	2,942	2,922
2	2	1	0,533	0,505	3	2,632	2,612
3	2	1	0,220	0,192	3	2,393	2,373
4	2	1	-0,053	-0,082	3	2,192	2,172
5	2	1	-0,300	-0,328	3	2,015	1,995
1	3	1	0,862	0,807	3	2,902	2,866
2	3	1	0,475	0,421	3	2,592	2,556
3	3	1	0,162	0,107	3	2,353	2,317
4	3	1	-0,111	-0,166	3	2,152	2,116
5	3	1	-0,358	-0,413	3	1,975	1,939
1	4	1	0,815	0,736	3	2,871	2,821
2	4	1	0,428	0,349	3	2,560	2,510
3	4	1	0,115	0,036	3	2,322	2,272
4	4	1	-0,158	-0,237	3	2,121	2,071
5	4	1	-0,405	-0,484	3	1,944	1,894
2	5	1	0,387	0,286	3	2,534	2,471
3	5	1	0,074	-0,027	3	2,296	2,233
4	5	1	-0,199	-0,300	3	2,095	2,032
5	5	1	-0,446	-0,547	3	1,918	1,855
2	6	1	0,351	0,228	3	2,511	2,437
3	6	1	0,038	-0,085	3	2,273	2,198
4	6	1	-0,235	-0,358	3	2,072	1,997
5	6	1	-0,482	-0,605	3	1,895	1,820
1	7	1	0,706	0,562	3	2,802	2,716
2	7	1	0,319	0,176	3	2,491	2,405
3	7	1	0,006	-0,138	3	2,253	2,167
4	7	1	-0,268	-0,411	3	2,052	1,966
5	7	1	-0,514	-0,658	3	1,875	1,789

Forrás: saját szerkesztés

Megfigyelhető, hogy az elégtelen jegyek esetében ez a metrika negatív értékeket is felvehetett. Minimuma a TVSZ-ben megengedett maximális próbálkozásszámmal vizsgajegyves tárgynál -0,235, gyakorlati jegynél pedig -0,358. Ennél kisebb értéket csak szélsőséges esetekben vehet fel.

Mindhárom metrika kialakításánál külön figyelmet fordítottunk azokra az esetekre, amikor a hallgató felvette a kurzust, de nem kapott végső jegyet belőle. Ez gyakorlati jegyes tárgyaknál azt jelenti, hogy nem teljesítette szorgalmi időszakban a jegyszerzés feltételét (például a megengedett-

nél többször hiányzott a tanóráról), vizsgajegyes tárgyaknál pedig azt, hogy nem ment el vizsgázni. Ilyenkor a végső bejegyzés 0-ra lett kódolva. Ebben az esetben a modellek jobb metrika értéket adnának annak a hallgatónak, aki nem ment el vizsgázni, azzal szemben, aki egyszer megpróbálta. Így ez azzal ekvivalens, mintha a hallgató fel sem vette volna a tárgyat, hiszen semmiféle nyoma nincs annak, hogy bármiféle energiát fektetett volna az adott kurzusba. Ezt a külön kitélt az Excel HA függvényének segítségével oldottuk meg.

7.2.6. A három metrika vizsgálata az IRF és ÜI kurzusok eredményein

Példaként vizsgáljuk meg, hogy az Informatikai rendszerek fejlesztése és az Üzleti intelligencia tárgyaknál hogyan alakulnak a kialakított metrikákon a hallgatói eredmények. Mivel csak azokra a tanulókra végezzük el a vizsgálatot, akik felvették a tárgyat, így az eredmények adatbázistól függetlenek. Mindkét tárgy gyakorlati jegyes, így a vizsgajegyeseknél lényegesen kevesebb kimenet lehetséges a metrika értékekre, és a hallgatók jellemzően kevesebb próbálkozással megszerezték a végső érdemjegyeiket.

Az IRF-nél a vizsgált adathalmazon előforduló tárgyfelvétel, vizsga és jegy kombinációk és a hozzájuk tartozó metrika értékek a 16. táblázatban láthatók. Megfigyelhető, hogy negatív metrika érték csak az a) és c) esetében, 0-s bejegyzésnél fordult elő, illetve, hogy a 0 vizsgázás esetén (azaz, ha a hallgató csak „Nem vizsgázott”² és „Megtagadva” bejegyzéseket szerzett) a metrikák egységesen 0 értéket vesznek fel. Láthatunk arra is példát, hogy valaki egy tárgyfelvétel mellett kétszer is vizsgázott, ennek egyik oka lehetett az, hogy szorgalmi időszak végén beírták a gyakorlati jegyét (elégtelent), de vizsgaidőszakban javítási lehetőséget kapott például pótló zárthelyi dolgozat formájában, ha valami miatt azt nem tudta megírni félév közben.

² Gyakorlati jegyes tárgy esetében is előfordult, hogy a hallgatók végső bejegyzése „Nem vizsgázott” lett.

16. táblázat: IRF-ből elért eredmények és metrika értékeik

Tárgyfelvétel	Vizsga	Jegy	a) metrika	b) metrika	c) metrika
1	0	0	0,000	0,000	0,000
3	2	0	-0,653	0,000	-1,011
1	1	1	1,000	1,000	1,000
2	1	1	0,689	0,918	0,613
2	2	1	0,586	0,881	0,505
3	3	1	0,268	0,804	0,107
1	1	2	2,000	2,000	2,000
1	2	2	1,896	1,920	1,922
2	1	2	1,689	1,827	1,689
2	2	2	1,586	1,747	1,612
3	3	2	1,268	1,569	1,317
1	1	3	3,000	3,000	3,000
2	1	3	2,689	2,725	2,689
2	2	3	2,586	2,597	2,612
3	3	3	2,268	2,287	2,317
1	1	4	4,000	4,000	4,000
2	2	4	3,586	3,428	3,505
1	1	5	5,000	5,000	5,000

Forrás: saját szerkesztés

A 17. táblázatban az Üzleti intelligencia metrika értékeiből látszik, hogy a vizsgált évfolyamokban a hallgatók 1-2 próbálkozással teljesíteni tudták a tárgyat, így csak 9 különböző eredményt értek el. Itt is egy esetben előfordult a vizsgaidőszaki javításnak köszönhetően a tárgyfelvételnél magasabb vizsgaszám. Láthatjuk, hogy 2 vizsga vagy tárgyfelvétel esetében még nincs nagy eltérés a metrika értékek között. Két tárgyfelvétellel és vizsgával elért közepes érdemjegy az a) metrikánál 2,586, a b) metrikánál 2,597, a c)-nél pedig 2,612 értéknek felelt meg.

17. táblázat: ÜI-ből elért eredmények és metrika értékeik

Tárgyfelvétel	Vizsga	Jegy	a) metrika	b) metrika	c) metrika
1	1	1	1,000	1,000	1,000
2	2	1	0,586	0,881	0,505
1	1	2	2,000	2,000	2,000
1	2	2	1,896	1,920	1,922
1	1	3	3,000	3,000	3,000
2	1	3	2,689	2,725	2,689
2	2	3	2,586	2,597	2,612
1	1	4	4,000	4,000	4,000
1	1	5	5,000	5,000	5,000

Forrás: saját szerkesztés

7.3. RANGKORRELÁCIÓ

Az elemzés következő szakaszában a tantárgyak közötti kapcsolatok vizsgálatát helyeztük fókuszba, az előző pontokban részletezett három metrika értékeit megvizsgálva a három adatbázison. Továbbá megfigyeltük ezek viselkedését és jóságát azzal a céllal, hogy a későbbiekben már csak egy kiválasztott adatbázis és metrika legyen az, amivel dolgozunk.

7.3.1. Rangkorrelációs mátrix kialakítása

Az érdemjegyek alapján képzett metrika értékek sorrendi skálán értelmezett változók, ezért a hagyományos korrelációs együttható helyett ezen attribútumok közötti kapcsolatok szorosságára a Spearman-féle rangkorrelációs együttható számítható. Ez, ahogyan arra a 4.1.4.1. pontban részletesen kitértünk, a változók értékeiből rangsorokat képez, majd az így kapott helyezések egymáshoz viszonyított helyzete alapján értelmezi a két attribútum közötti kapcsolat szorosságát (Zar, 2005).

A korrelációs értékeket a tizenhárom változóra rangkorrelációs mátrixok segítségével ábrázoltuk mindhárom metrika esetében, melyeknek oszlop és sor fejlécei az egyes tantárgyak. A mátrix elemei pedig az adott sor-oszlop fejléceihez tartozó tantárgyak metrika értékei közötti rangkorrelációs együtthatót tartalmazzák. Ezen gondolatmenet alapján alakítottunk ki a RStudio segítségével kilenc rangkorrelációs mátrixot, hogy a három adatbázis és három modell összes kombinációját vizsgálni tudjuk és összevegyük, milyen erősségű kapcsolatok tapasztalhatók a különböző esetekben. Az *Összes null* adatbázison alkalmazott, c) metrika értékeire elvégzett vizsgálat eredményeképpen kapott mátrix a 18. táblázatban látható:

18. táblázat: Rangkorrelációs mátrix a c) metrika értékeire az Összes null adatbázison

Rangkorreláció az Összes null adatbázis c) metrika értékeire													
	ui	szfl	szfl	szamt	anal	lin	opk	adatb	infa	irf	valsz	statl	statll
ui	1,0000	0,2776	0,3696	0,3530	0,3433	0,3053	0,4432	0,4372	0,2715	0,5207	0,3119	0,4103	0,5097
szfl	0,2776	1,0000	0,4730	0,4132	0,3804	0,3470	0,4760	0,3853	0,4492	0,3246	0,3718	0,4516	0,4338
szfl	0,3696	0,4730	1,0000	0,3244	0,3291	0,3172	0,4522	0,4319	0,3892	0,4076	0,3623	0,3818	0,5511
szamt	0,3530	0,4132	0,3244	1,0000	0,5923	0,5844	0,5056	0,3933	0,4442	0,3655	0,4974	0,5873	0,5246
anal	0,3433	0,3804	0,3291	0,5923	1,0000	0,5762	0,4910	0,3638	0,4824	0,3979	0,4554	0,6542	0,4985
lin	0,3053	0,3470	0,3172	0,5844	0,5762	1,0000	0,4855	0,4390	0,4845	0,3861	0,4941	0,5562	0,4365
opk	0,4432	0,4760	0,4522	0,5056	0,4910	0,4855	1,0000	0,5043	0,3911	0,5047	0,4863	0,5477	0,6620
adatb	0,4372	0,3853	0,4319	0,3933	0,3638	0,4390	0,5043	1,0000	0,3669	0,5214	0,4282	0,4453	0,5426
infa	0,2715	0,4492	0,3892	0,4442	0,4824	0,4845	0,3911	0,3669	1,0000	0,3543	0,3516	0,5516	0,3900
irf	0,5207	0,3246	0,4076	0,3655	0,3979	0,3861	0,5047	0,5214	0,3543	1,0000	0,3237	0,3967	0,5079
valsz	0,3119	0,3718	0,3623	0,4974	0,4554	0,4941	0,4863	0,4282	0,3516	0,3237	1,0000	0,4998	0,4740
statl	0,4103	0,4516	0,3818	0,5873	0,6542	0,5562	0,5477	0,4453	0,5516	0,3967	0,4998	1,0000	0,6099
statll	0,5097	0,4338	0,5511	0,5246	0,4985	0,4365	0,6620	0,5426	0,3900	0,5079	0,4740	0,6099	1,0000

Színmagyarázat: világossárga: $0 < \rho_{on} < 0,3$; sárga: $0,3 \leq \rho_{on} < 0,5$; világoskék: $0,5 \leq \rho_{on} < 0,7$; sötétkék $\rho_{on} \geq 0,7$

Forrás: saját szerkesztés

A szorosság mértékét mutató együtthatók színezése a hagyományostól eltérő módon valósult meg, a 0,3 és 0,7 vágópontokon kívül – melyek a gyenge-közepes és a közepes-erős kapcsolatokat választják el –, a 0,5 értékénél egy új vágópontot hoztunk létre. Ennek oka, hogy a mátrixban szereplő korrelációs együtthatók nagy része 0,3 és 0,7 közé esett, azonban mi ezek között is szeretnénk volna látni az eltéréseket, így ezt az új színezési módot alkalmaztuk az összes mátrix esetében.

7.3.2. Rangkorreláció vizsgálata a három adatbázisra

A továbbiakban a rangkorrelációs vizsgálatokat mindhárom adatbázisra elvégeztük, ezek eredményeit foglaljuk össze az alábbiakban:

7.3.2.1. Az Összes adatbázis:

Ahogy azt már a 6.3.4. pontban is említettük, az Összes adatbázis úgy lett kialakítva, hogy abban az esetben, ha a hallgató nem vette fel az adott tárgyat, az ahhoz tartozó metrika értéke 0 lett.

19. táblázat: Rangkorrelációs mátrix a c) metrika értékeire az Összes adatbázison

Rangkorreláció az Összes adatbázis c) metrika értékeire													
	ui	szfl	szflI	szamt	anal	lin	opk	adatb	infa	irf	valsz	statI	statII
ui	1,0000	0,5257	0,6473	0,5902	0,5230	0,5926	0,6809	0,6941	0,4460	0,7410	0,5722	0,5662	0,7114
szfl	0,5257	1,0000	0,6860	0,5744	0,4695	0,5360	0,6183	0,5746	0,5574	0,5819	0,5715	0,5383	0,6040
szflI	0,6473	0,6860	1,0000	0,5731	0,5019	0,5693	0,6536	0,6545	0,5430	0,6811	0,6073	0,5285	0,7093
szamt	0,5902	0,5744	0,5731	1,0000	0,5497	0,6136	0,6157	0,5773	0,4924	0,5839	0,6113	0,5914	0,6400
anal	0,5230	0,4695	0,5019	0,5497	1,0000	0,5836	0,5762	0,4864	0,4595	0,5439	0,5247	0,6300	0,5967
lin	0,5926	0,5360	0,5693	0,6136	0,5836	1,0000	0,6236	0,6078	0,5368	0,6094	0,6450	0,5764	0,5971
opk	0,6809	0,6183	0,6536	0,6157	0,5762	0,6236	1,0000	0,6564	0,5054	0,6892	0,6232	0,6434	0,7905
adatb	0,6941	0,5746	0,6545	0,5773	0,4864	0,6078	0,6564	1,0000	0,4872	0,7234	0,6060	0,5466	0,6862
infa	0,4460	0,5574	0,5430	0,4924	0,4595	0,5368	0,5054	0,4872	1,0000	0,4942	0,4901	0,5508	0,4941
irf	0,7410	0,5819	0,6811	0,5839	0,5439	0,6094	0,6892	0,7234	0,4942	1,0000	0,5829	0,5481	0,6972
valsz	0,5722	0,5715	0,6073	0,6113	0,5247	0,6450	0,6232	0,6060	0,4901	0,5829	1,0000	0,5693	0,6309
statI	0,5662	0,5383	0,5285	0,5914	0,6300	0,5764	0,6434	0,5466	0,5508	0,5481	0,5693	1,0000	0,6922
statII	0,7114	0,6040	0,7093	0,6400	0,5967	0,5971	0,7905	0,6862	0,4941	0,6972	0,6309	0,6922	1,0000

Színmagyarázat: sárga: $0,3 \leq \rho_0 < 0,5$; világoskék: $0,5 \leq \rho_0 < 0,7$; sötétkék $\rho_0 \geq 0,7$

Forrás: saját szerkesztés

Ezen adatbázis metrika értékeire elvégzett rangkorrelációs vizsgálatok eredményei a c) metrika esetében jelentős, közepesen szoros és szoros kapcsolatokat mutatnak (lásd 19. táblázat). Látható, hogy ebben az esetben a legkisebb együttható – 0,4460 az Informatika alapjai és az Üzleti intelligencia eredményei között – esetében is közepesen erős kapcsolat figyelhető meg a tárgyak között. A legerősebb kapcsolat kiemelkedően magas, közel 0,8 az Operációkutatás és a Statisztika II. tantárgyakból elért eredmények között. A rangkorrelációs mátrixokat a másik két metrika értékeire is elkészítettük és ezeket összehasonlítva megállapítható, hogy csak kis eltérések vannak a

korrelációs együtthatók között, ami abból következik, hogy a metrikák közötti jelentősebb eltérések csak a szélsőségesebb értékeknél jelentek meg, viszont ezekből a vizsgált adathalmazon kevesebb volt. Az *a)* és *c)* metrika értékei között 0,003-nál nagyobb eltérés nincs, a *b)* metrika esetében ez kicsivel nagyobb mértékű, esetenként a 0,01-et is eléri, azonban meg kell jegyeznünk, hogy ez a kapcsolatok mértékéhez képest nem jelentős.

7.3.2.2. Az Összes null adatbázis

Az *Összes null* adatbázis – amely a fel nem vett tárgyaknál NULL metrika értéket rendelt a hallgatók eredményéhez – alapján kialakított rangkorrelációs mátrixok esetében sem figyelhetők meg különösebb eltérések a három metrikára elvégzett vizsgálatok között, az *a)* és *c)* modellek közti legnagyobb eltérés 0,009, az *a)* és *b)* között pedig 0,01. A 18. táblázat jól mutatja, hogy ennél az adatbázisnál a tantárgyak között legjellemzőbb a közepesen erős kapcsolat, a három metrikára vonatkozó rangkorrelációs együtthatók átlagosan 0,442 és 0,446 között alakultak, 0,09 szórással.

7.3.2.3. A Szűrt adatbázis

A *Szűrt* adatbázis adatain kialakított rangkorrelációs mátrixokban a másik két adatbázishoz képest még kevésbé figyelhető meg eltérés a három metrikára vonatkozó korrelációs együtthatók között. Az *a)* és *c)* metrika értékei minden esetben megegyeznek, az *a)* és *b)* közötti különbség is csak negyedik tizedesjegynél érzékelhető. Ennek oka, hogy az adatbázis kialakításakor mindazon hallgatók adatait, akik legalább egy tárgyat nem vettek fel, eltávolítottuk az adathalmazból. Mivel megfigyelhető egy olyan tendencia, hogy aki egy tárgyat nem teljesít, annak a többi kurzussal is nehézségei adódnak, ezért pont a leggyengébb hallgatók eredményeit szűrtük ki. Azaz pont azokat az adatokat távolítottuk el, amelyeknek az alapértelmezett jegyeiktől eltérő metrika értékei vannak, tehát az ezek közötti különbségeket csökkentettük. A mátrixban megfigyelhető, hogy a legjellemzőbb eset itt is a közepesen erős kapcsolat, de valamivel gyengébb mértékű – átlagosan 0,391, 0,08 szórással –, mint az előző két adatbázison, és itt már nagyobb számban találhatók gyenge kapcsolatok is a tárgyak között. A *c)* metrika értékeire vonatkozó mátrix a 20. táblázatban látható.

20. táblázat: Rangkorrelációs mátrix a c) metrika értékeire a Szűrt adatbázison

Rangkorreláció a Szűrt adatbázis c) metrika értékeire													
	ui	szfl	szflI	szamt	anal	lin	opk	adatb	infa	irf	valsz	statI	statII
ui	1,0000	0,2430	0,3555	0,3365	0,3418	0,2746	0,4309	0,4215	0,2566	0,5203	0,2723	0,3922	0,4960
szfl	0,2430	1,0000	0,4525	0,3489	0,3219	0,2828	0,4226	0,3005	0,4326	0,3026	0,2817	0,4061	0,3886
szflI	0,3555	0,4525	1,0000	0,3136	0,2780	0,2776	0,3761	0,3494	0,3882	0,3782	0,3055	0,3404	0,5083
szamt	0,3365	0,3489	0,3136	1,0000	0,5107	0,4399	0,4631	0,3185	0,3840	0,3391	0,4567	0,4909	0,4940
anal	0,3418	0,3219	0,2780	0,5107	1,0000	0,4500	0,4328	0,2714	0,3660	0,3713	0,3776	0,5500	0,4574
lin	0,2746	0,2828	0,2776	0,4399	0,4500	1,0000	0,4377	0,3398	0,4185	0,3388	0,4278	0,4615	0,3572
opk	0,4309	0,4226	0,3761	0,4631	0,4328	0,4377	1,0000	0,3930	0,3790	0,4685	0,4241	0,4793	0,5969
adatb	0,4215	0,3005	0,3494	0,3185	0,2714	0,3398	0,3930	1,0000	0,3418	0,4628	0,3328	0,3602	0,4598
infa	0,2566	0,4326	0,3882	0,3840	0,3660	0,4185	0,3790	0,3418	1,0000	0,3498	0,2996	0,4523	0,3888
irf	0,5203	0,3026	0,3782	0,3391	0,3713	0,3388	0,4685	0,4628	0,3498	1,0000	0,2708	0,3469	0,4782
valsz	0,2723	0,2817	0,3055	0,4567	0,3776	0,4278	0,4241	0,3328	0,2996	0,2708	1,0000	0,4194	0,4034
statI	0,3922	0,4061	0,3404	0,4909	0,5500	0,4615	0,4793	0,3602	0,4523	0,3469	0,4194	1,0000	0,5643
statII	0,4960	0,3886	0,5083	0,4940	0,4574	0,3572	0,5969	0,4598	0,3888	0,4782	0,4034	0,5643	1,0000

Színmagyarázat: világossárga: $0 < \rho_{sz} < 0,3$; sárga: $0,3 \leq \rho_{sz} < 0,5$; világoskék: $0,5 \leq \rho_{sz} < 0,7$;

sötétkék $\rho_{sz} \geq 0,7$

Forrás: saját szerkesztés

7.3.3. A rangkorrelációs mátrixok összehasonlítása a három adatbázis esetében

Ezt követően megvizsgáltuk a három adatbázis metrika értékei alapján kialakított rangkorrelációs mátrixok eltéréseit, hogy lássuk, milyen mértékű eltéréseket okoz a tárgyak közötti kapcsolatok erősségében az, hogy a fel nem vett tárgyak eredményeit különböző módokon kezeltük.

7.3.3.1. Összes és Szűrt közötti eltérések:

A két adatbázis alapján kapott rangkorrelációs együtthatók összevetéséhez létrehoztuk mindhárom metrikára vonatkozóan a rangkorrelációs mátrixok különbségmátrixát. Mivel az Összes adatbázis több egyedet tartalmaz és az ebből kialakított mátrixban szorosabb kapcsolatok vannak, ezért ebből vontuk ki a Szűrt adatbázishoz tartozó mátrixot.

**21. táblázat: Összes és Szűrt rangkorrelációs mátrixok különbsége a
c) metrika értékein**

Összes és Szűrt rangkorrelációs mátrixok különbsége													
	ui	szfl	szflI	szamt	anal	lin	opk	adatb	infa	irf	valsz	statI	statII
ui	0,0000	0,2827	0,2918	0,2537	0,1812	0,3180	0,2500	0,2726	0,1894	0,2208	0,2999	0,1741	0,2154
szfl	0,2827	0,0000	0,2335	0,2255	0,1476	0,2531	0,1957	0,2741	0,1247	0,2793	0,2899	0,1322	0,2154
szflI	0,2918	0,2335	0,0000	0,2594	0,2239	0,2916	0,2775	0,3050	0,1548	0,3028	0,3017	0,1881	0,2010
szamt	0,2537	0,2255	0,2594	0,0000	0,0391	0,1738	0,1526	0,2588	0,1084	0,2448	0,1546	0,1005	0,1460
anal	0,1812	0,1476	0,2239	0,0391	0,0000	0,1336	0,1434	0,2150	0,0935	0,1726	0,1471	0,0800	0,1394
lin	0,3180	0,2531	0,2916	0,1738	0,1336	0,0000	0,1858	0,2679	0,1183	0,2705	0,2172	0,1149	0,2399
opk	0,2500	0,1957	0,2775	0,1526	0,1434	0,1858	0,0000	0,2635	0,1265	0,2207	0,1990	0,1641	0,1936
adatb	0,2726	0,2741	0,3050	0,2588	0,2150	0,2679	0,2635	0,0000	0,1454	0,2606	0,2732	0,1864	0,2264
infa	0,1894	0,1247	0,1548	0,1084	0,0935	0,1183	0,1265	0,1454	0,0000	0,1444	0,1905	0,0985	0,1053
irf	0,2208	0,2793	0,3028	0,2448	0,1726	0,2705	0,2207	0,2606	0,1444	0,0000	0,3121	0,2013	0,2190
valsz	0,2999	0,2899	0,3017	0,1546	0,1471	0,2172	0,1990	0,2732	0,1905	0,3121	0,0000	0,1499	0,2275
statI	0,1741	0,1322	0,1881	0,1005	0,0800	0,1149	0,1641	0,1864	0,0985	0,2013	0,1499	0,0000	0,1279
statII	0,2154	0,2154	0,2010	0,1460	0,1394	0,2399	0,1936	0,2264	0,1053	0,2190	0,2275	0,1279	0,0000

Színmagyarázat: világossárga: $0 < \rho_o - \rho_{sz} < 0,3$; sárga: $0,3 \leq \rho_o - \rho_{sz} < 0,5$

Forrás: saját szerkesztés

Ennek eredményeképpen előálltak tantárgypáronként az *Összes* és a *Szűrt* adatbázis alapján kapott korrelációs együtthatók értékeinek különbségei. Az eltérések annak köszönhetők, hogy míg az *Összes* adatbázisban a fel nem vett tárgyak esetén a metrika 0 értéket vesz fel, addig a *Szűrt* adatbázisban ezek az egyedek teljesen el lettek távolítva az adathalmazból. A 21. táblázatban látható, hogy a *c) metrikát* alapul vevő adatbázisoknál az eltérések jelentősek, több esetben a 0,3-et is meghaladják, ami már közepesen gyenge kapcsolatnak felel meg, a különbségek átlaga 0,202. Az *a)* és *b) metrika* értékeire is hasonló mértékűek az eltérések, ezek átlagai rendre 0,201 és 0,205.

7.3.3.2. *Összes null és Szűrt közötti eltérések:*

Az *Összes null* és a *Szűrt* adatbázisok alapján kialakított rangkorrelációs mátrixok különbségmátrixa a 22. táblázatban látható. Ebben az esetben az eltérések csak minimálisak voltak, átlagosan 0,05 mindhárom metrika esetében. A legnagyobb eltérés 0,1445, a Lineáris algebra és Számítástudomány kapcsolatának szorossága között figyelhető meg.

22. táblázat: Összes null és Szűrt rangkorrelációs mátrixok különbsége

a c) metrika értékein

Összes null és Szűrt rangkorrelációs mátrixok különbsége													
	ui	szfl	szfil	szamt	anal	lin	opk	adatb	infa	irf	valsz	statl	statll
ui	0,0000	0,0345	0,0141	0,0165	0,0015	0,0307	0,0123	0,0157	0,0149	0,0004	0,0396	0,0182	0,0136
szfl	0,0345	0,0000	0,0205	0,0643	0,0585	0,0642	0,0534	0,0848	0,0166	0,0220	0,0901	0,0455	0,0452
szfil	0,0141	0,0205	0,0000	0,0107	0,0511	0,0396	0,0761	0,0825	0,0010	0,0293	0,0568	0,0414	0,0427
szamt	0,0165	0,0643	0,0107	0,0000	0,0816	0,1445	0,0425	0,0748	0,0602	0,0265	0,0407	0,0964	0,0306
anal	0,0015	0,0585	0,0511	0,0816	0,0000	0,1263	0,0582	0,0923	0,1164	0,0265	0,0778	0,1042	0,0411
lin	0,0307	0,0642	0,0396	0,1445	0,1263	0,0000	0,0477	0,0992	0,0660	0,0473	0,0663	0,0947	0,0793
opk	0,0123	0,0534	0,0761	0,0425	0,0582	0,0477	0,0000	0,1114	0,0121	0,0363	0,0622	0,0684	0,0651
adatb	0,0157	0,0848	0,0825	0,0748	0,0923	0,0992	0,1114	0,0000	0,0251	0,0586	0,0954	0,0851	0,0827
infa	0,0149	0,0166	0,0010	0,0602	0,1164	0,0660	0,0121	0,0251	0,0000	0,0045	0,0520	0,0993	0,0011
irf	0,0004	0,0220	0,0293	0,0265	0,0265	0,0473	0,0363	0,0586	0,0045	0,0000	0,0529	0,0498	0,0298
valsz	0,0396	0,0901	0,0568	0,0407	0,0778	0,0663	0,0622	0,0954	0,0520	0,0529	0,0000	0,0804	0,0706
statl	0,0182	0,0455	0,0414	0,0964	0,1042	0,0947	0,0684	0,0851	0,0993	0,0498	0,0804	0,0000	0,0456
statll	0,0136	0,0452	0,0427	0,0306	0,0411	0,0793	0,0651	0,0827	0,0011	0,0298	0,0706	0,0456	0,0000

Színmagyarázat: világossárga: $0 < \rho_{on} - \rho_{sz} < 0,3$

Forrás: saját szerkesztés

7.3.3.3. Összes és Összes null közötti eltérések:

Végül az *Összes* és az *Összes null* adatbázisok alapján kialakított mátrixok különbségét vizsgáltuk meg. Itt már nagyobb eltérések láthatók, a maximális érték a c) metrika esetében 0,2873. Az átlagos eltérés mértéke a három metrikára 0,148, 0,149 és 0,150. Megfigyelhető, hogy ebben az esetben a kapott különbségmátrix negatív értékeket is tartalmaz. A két adatbázis közötti különbség az volt, hogy míg az *Összes* esetén az egy tárgyat fel sem vevő hallgatók metrika értéke az adott tárgyból 0 lett, addig az *Összes null* adatbázison NULL értéket kaptak. Így a negatív értékek azt jelentik, hogy voltak olyan tantárgypárok, amelyek között erősebb kapcsolatot okozott a metrika érték hiánya, mint annak 0-val való pótlása. Az eltérések a c) metrika esetében a 23. táblázatban láthatók.

23. táblázat: Összes és Összes null rangkorrelációs mátrixok különbsége

a c) metrika értékein

Összes és Összes null rangkorrelációs mátrixok különbsége													
	ui	szfl	szfil	szamt	anal	lin	opk	adatb	infa	irf	valsz	statl	statll
ui	0,0000	0,2481	0,2777	0,2373	0,1797	0,2873	0,2377	0,2569	0,1745	0,2203	0,2603	0,1559	0,2018
szfl	0,2481	0,0000	0,2130	0,1612	0,0891	0,1890	0,1423	0,1893	0,1081	0,2573	0,1998	0,0867	0,1702
szfil	0,2777	0,2130	0,0000	0,2487	0,1728	0,2521	0,2013	0,2226	0,1538	0,2735	0,2450	0,1466	0,1582
szamt	0,2373	0,1612	0,2487	0,0000	-0,0425	0,0292	0,1101	0,1840	0,0482	0,2184	0,1139	0,0041	0,1154
anal	0,1797	0,0891	0,1728	-0,0425	0,0000	0,0073	0,0851	0,1226	-0,0228	0,1461	0,0693	-0,0242	0,0982
lin	0,2873	0,1890	0,2521	0,0292	0,0073	0,0000	0,1381	0,1687	0,0523	0,2233	0,1509	0,0202	0,1606
opk	0,2377	0,1423	0,2013	0,1101	0,0851	0,1381	0,0000	0,1521	0,1143	0,1845	0,1368	0,0957	0,1285
adatb	0,2569	0,1893	0,2226	0,1840	0,1226	0,1687	0,1521	0,0000	0,1203	0,2020	0,1778	0,1013	0,1436
infa	0,1745	0,1081	0,1538	0,0482	-0,0228	0,0523	0,1143	0,1203	0,0000	0,1399	0,1385	-0,0008	0,1042
irf	0,2203	0,2573	0,2735	0,2184	0,1461	0,2233	0,1845	0,2020	0,1399	0,0000	0,2591	0,1515	0,1893
valsz	0,2603	0,1998	0,2450	0,1139	0,0693	0,1509	0,1368	0,1778	0,1385	0,2591	0,0000	0,0695	0,1569
statl	0,1559	0,0867	0,1466	0,0041	-0,0242	0,0202	0,0957	0,1013	-0,0008	0,1515	0,0695	0,0000	0,0823
statll	0,2018	0,1702	0,1582	0,1154	0,0982	0,1606	0,1285	0,1436	0,1042	0,1893	0,1569	0,0823	0,0000

Színmagyarázat: világossárga: $0 < \rho_o - \rho_{on} < 0,3$; piros: $\rho_o - \rho_{on} < 0$

Forrás: saját szerkesztés

7.3.4. A rangkorrelációs eredmények vizsgálata az IRF és ÜI kurzusokon

Az elemzés ezen szakaszában is megvizsgáltuk, hogy a kapott eredmények mit jelentenek az Informatikai rendszerek fejlesztése és az Üzleti intelligencia tantárgyakra vonatkozóan. Láthattuk, hogy a három metrika szempontjából csak kis eltérések figyelhetők meg a rangkorrelációs együtt-hatók között, ezért a könnyebb érthetőség kedvéért csak a korábbiakban is példaként vett *c) metrika* értékei közötti kapcsolatokat vizsgáljuk meg. Az *Összes* adatbázis rangkorrelációs mátrixát elemezve láthatjuk, hogy az IRF-et tekintve a legerősebb kapcsolat az ÜI-vel és az Adatb-vel figyel-hető meg, mindkét esetben 0,7 feletti a kapcsolat szorossága. Ezekben az esetekben a legnagyobb mértékű az az együttmozgás, hogyha az egyik tárgyból rosszul teljesített a hallgató, akkor a másik tárgyból is nehézségei voltak. Magas értékeket találhatunk még az IRF esetében a StatII (0,6972), az Opk (0,6892) és a SzfII (0,6811) tárgyakkal vett rangkorrelációknál. Leggyengébb, kicsivel 0,5 alatti az Infa és az IRF közötti kapcsolat. A *Szűrt* adatbázison kevésbé szoros kapcsolatok figyel-hetők meg. A legerősebb kapcsolat ebben az esetben is az ÜI-ből elért eredményekkel van (0,5203), a többi érték 0,5 alatti. Egy gyenge kapcsolatot láthatunk, a Valószínűségyszámítással (0,2708), vi-szont meglepő módon, az *Összes* adatbázison ez a kapcsolat 0,5829-es értékkel a többi tárgyhöz képest átlagosnak számított. Emellett itt, a korábbtól eltérő módon az Informatikával lévő kapcsola-t is erősebbnek bizonyult (0,3498). A két mátrix közötti eltérés oka tehát, hogy míg az *Összes* adatbázis 0-nak kódolja a fel sem vett tárgyakat, addig a *Szűrtből* ezek a hallgatók el is lettek távo-lítva. Ennek hatására a kapcsolatok szorossága nöhetett vagy csökkenhetett az egyes adatbáziso- kon. A Valószínűségyszámításnál lévő eltérés az *Összes* és a *Szűrt* korrelációs mátrixoknál is szem- betűnő, ebben az esetben van a legnagyobb (0,3121) eltérés a két rangkorrelációs együtt-ható között. Ez alapján érdemes még a SzfII értékeit is kiemelni, ami az *Összes* adatbázison 0,6811, a *Szűrt* adatbázison pedig 0,3882 szorosságú lett. Az *Összes null* adatbázison alapuló rangkorrelációs ér- tékek itt, az IRF esetében is a legtöbbször a másik kettő közötti értékeket vette fel. Négy esetben volt 0,5 feletti kapcsolat, legmagasabb itt is az Adatb (0,5214) és az ÜI (0,5207) tárgyakból elért eredményekkel. Gyenge kapcsolat azonban ebben az esetben nem volt.

Az Üzleti intelligenciából elért eredmények közül az *Összes* adatbázison a legszorosabb kap- csolat a már korábban is említett IRF-fel figyelhető meg (0,7410), de szintén 0,7 feletti a StatII-vel (0,7114). Leggyengébb és egyben egyedüli 0,5 alatti itt is az Infa-ból elért eredményekkel a kap- csolat (0,4460), az átlagos érték 0,6076 volt. *Szűrt* adatbázison alapuló rangkorrelációs mátrixban figyelhető meg az összes kapcsolat közül a legkisebb: 0,2430 a SzfI tárgy eredményeivel. További

három tárggyal szintén 0,3 alatti a rangkorrelációs együttható értéke: a Lin (0,2746), az Infa (0,2566) és a Valsz (0,2723). E kurzusok közül csak a Lineáris algebra volt olyan, ahol a két adatbázis közötti érték kiemelkedően eltér: 0,318 a kettő különbsége. Itt is a többi tárggyhoz képest magas, 0,5 feletti volt a kapcsolat az IRF metrika értékeivel (0,5203). Az *Összes null* adatbázis az ÜI esetében is a másik kettő között képez átmenetet. Ezen adatbázis értékek között kizárólag itt található gyenge kapcsolat: 0,2776 a Szfl és 0,2715 az Infa esetében. Az *Összes* adatbázishoz hasonlóan a két legmagasabb érték itt is az IRF (0,5207) és a StatII (0,5097) eredményeivel lett.

7.3.5. Döntés a továbbiakban használt adatbázis és metrika mellett

A három adatbázis rangkorrelációs mátrixait és ezek különbségmátrixait vizsgálva amellet döntöttünk, hogy az *Összes null* adatbázis lesz az, amit kiválasztunk és ezután már csak ezzel haladunk tovább az elemzésekkel. A rangkorrelációs kapcsolatok mellett a döntést segítette az, hogy a *Szűrt* adatbázisból lényeges, értékes adatok lettek eltávolítva azzal, hogy véglegesen töröltük azoknak a hallgatóknak az összes eredményét, akik legalább egy tárgyat nem vettek fel. Az *Összes* adatbázis pedig a 0 értékek adásával felfelé torzítja a tárgyak közti kapcsolatokat. Ez a különbségmátrixon látszik: az *Összes null* és a *Szűrt* különbségmátrixában megfigyelhető kis különbségekből és az *Összes* és *Összes null* rangkorrelációs együtthatóinak nagyobb mértékű eltéréséből adódik, hogy a metrika értékek NULL-ra állításához képest a hallgatók értékeinek teljes eltávolítása kisebb mértékben torzítja a kapcsolatok szorosságának mértékét, mintha a metrika helyére 0 értéket rendelnénk. A kiválasztott *Összes null* adatbázis az, amely a rangkorrelációs együtthatók tekintetében a két szélsőséges eset között átmenetet képez.

Láthattuk, hogy a rangkorrelációs értékeknél, azaz a különböző tárgyakból elért eredmények közötti kapcsolatok szorosságánál csak minimális eltérés van a metrikák viselkedése között. Illetve megfigyelhettük, hogy az *a)* és *c)* *metrikák* a legtöbb esetben megegyeznek, míg a *b)* *metrika* többször vesz fel eltérő értékeket a rangkorrelációnál. Ezért úgy döntöttünk, hogy a három metrika közül a továbbiakban csak a *c)* *metrikával* foglalkozunk, mivel ez az, ami a későbbiekben használatos LAD regresszióhoz szükséges legszélesebb értékkészletet biztosítja. Emellett nagy előnye még, hogy az alapjait jelentő modellben sokkal könnyebben állíthatók a tárgyfelvelet és vizsgafelvelet büntetésének mértékei. Így, ha újabb vizsgálatok során eltérő kimeneteket keresünk a próbálkozások értékelésére, akkor ezt lehet legkönnyebben finomhangolni.

7.4. ELŐFELTÉTELEK VIZSGÁLATA

7.4.1. Előfeltételmátrix

A rangkorrelációs mátrixok alapján láthattuk, hogy bizonyos tantárgyak eredményei között szorosabb kapcsolat van. Szerettük volna kideríteni ennek okát, illetve azt is megvizsgálni, hogy mi állhat annak a hátterében, hogy a félévek során morzsolódnak le a hallgatók, mennyire van ebben szerepe annak, hogy az előkövetelményeket nem tudják teljesíteni és emiatt nem folytathatják tovább a tanulmányaikat.

A tárgyak felvételéhez kötelezően teljesítendő tantárgyaknak, azaz az előfeltételeknek a kapcsolatok szorosságára mért hatását vizsgálva érdemes szemügyre venni, hogy van-e valamilyen eltérés két olyan tárgy közötti korrelációnál, ahol az egyik előfeltétele a másiknak. Ehhez egy olyan, a 24. táblázatban látható mátrixot készítettünk, melyben 0 jelöli, ahol nem előfeltételes a kapcsolat, 1 pedig azt, ahol egyik előfeltétele a másiknak, majd ezt összevetettük a korábban elkészített rangkorrelációs mátrixokkal. A vizsgált 2011/2012, 2012/2013 és 2013/2014-es gazdaság-informatikus évfolyamoknál csak két előfeltételes kapcsolat volt a képzés operatív tantervében, így ez ugyanazt a három tárgyat érintette mindhárom évfolyam esetében. Ez megkönnyítette a vizsgálatot, hiszen így az évfolyamok eredményeit összesítve lehetett kezelni. Az előfeltételes kapcsolatok ezekben az években a Szoftver-technológia I. és II., illetve a Szoftver-technológia II. és az Informatikai rendszerek fejlesztése tárgyak között voltak.

24. táblázat: Előfeltételmátrix

Előfeltételmátrix														
	ui	szfl	szfil	szamt	anal	lin	opk	adatb	infa	irf	valsz	statl	statll	
ui	0	0	0	0	0	0	0	0	0	0	0	0	0	0
szfl	0	0	1	0	0	0	0	0	0	0	0	0	0	0
szfil	0	1	0	0	0	0	0	0	0	1	0	0	0	0
szamt	0	0	0	0	0	0	0	0	0	0	0	0	0	0
anal	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lin	0	0	0	0	0	0	0	0	0	0	0	0	0	0
opk	0	0	0	0	0	0	0	0	0	0	0	0	0	0
adatb	0	0	0	0	0	0	0	0	0	0	0	0	0	0
infa	0	0	0	0	0	0	0	0	0	0	0	0	0	0
irf	0	0	1	0	0	0	0	0	0	0	0	0	0	0
valsz	0	0	0	0	0	0	0	0	0	0	0	0	0	0
statl	0	0	0	0	0	0	0	0	0	0	0	0	0	0
statll	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Forrás: saját szerkesztés

Az így kapott előfeltétel változókat a rangkorrelációs együtthatókkal összevetve vizsgáltuk meg, hogy van-e kapcsolat a korreláció mértéke és aközött, hogy van-e előkövetelményes viszony a két tárgy között. Ehhez a vegyes kapcsolat mérésére szolgáló H mutatót számoltunk, melynek értéke 0,17 lett. Tehát elmondható, hogy a vizsgált adathalmazon csak gyenge kapcsolat van az előfeltétel léte és a korrelációs együttható értéke között. Természetesen ez annak is köszönhető, hogy a vizsgált kapcsolatoknak csak 2,56%-a volt előkövetelményes, egy több előfeltételt tartalmazó tantervi hálóban mindenképpen érdemes lehet elvégezni ezeket az elemzéseket.

7.4.2. Előfeltételes regresszió

Az előfeltételek további hatásainak feltárásához megvizsgáltuk, hogy az előkövetelmény tárgyból elért eredmény és az abból szerzett elégtelen jegy hatással van-e – és ha igen, milyen mértékben – arra, hogy a hallgató milyen eredményt ér el az arra a tantárgyra épülő kurzuson. Ennek vizsgálatára regressziós elemzési módszert alkalmaztunk. Ahogyan azt már a 4.1.4. pontban részletesen kifejtettük, a sorrendi skálán értelmezett metrika értékekre a hagyományos lineáris regresszió nem használható, viszont az érdemjegyek segítségével kialakított metrikával egy olyan széles érték-készletet kaptunk, amelyen már értelmezhető a medián és így a feltételes medián értékét becslő LAD regresszió alkalmazható. Az értékkészletek megfelelő szélességűek a medián értelmezéséhez, ezt bővebben a 7.5.1. pontban tárgyaljuk. Mivel a vizsgált adathalmazon egy háromszintes előfeltétel-hierarchia figyelhető meg (SzfII előfeltétele SzfI és IRF előkövetelménye SzfII), ezért az előfeltételeknek a hierarchia tetején álló IRF-re gyakorolt hatását elemeztük. A vizsgálat során egy olyan táblázatot alakítottunk ki, melyben az IRF tantárgy metrika értékei mellé a SzfI és SzfII tárgyak értékeit vettük fel. Ezek mellé továbbá egy dummy változót is létrehoztunk, amely azt mutatja, hogy az IRF előfeltételeiből volt-e legalább egy elégtelen bejegyzése az adott hallgatónak. Az így kapott adattáblára építettük fel a Gretl szoftver segítségével a LAD regressziós modelleket.

A többváltozós regressziós modell futtatása megadja, hogy a parciálisan értelmezhető regressziós együtthatók közül az előfeltételből való bukás értékének eggyel növekedése – azaz dummy változó lévén az, hogy valaki bukott – ceteris paribus mennyivel növeli meg a ráépülő tárgy metrika értékét az adott modellben.

7. ábra: LAD regresszió értékei

Model 1: LAD, using observations 1-463					
Dependent variable: irf_model1					
	coefficient	std. error	t-ratio	p-value	
const	0,596776	0,249520	2,392	0,0172	**
szfI_model1	0,000000	0,0721915	0,0000	1,0000	
szfII_model1	0,600806	0,0887721	6,768	4,00e-011	***
elofelt_bukas	-0,596776	0,232328	-2,569	0,0105	**
Median depend. var	3,000000	S.D. dependent var	1,245463		
Sum absolute resid	343,7059	Sum squared resid	415,6227		
Log-likelihood	-645,9807	Akaike criterion	1299,961		
Schwarz criterion	1316,512	Hannan-Quinn	1306,477		

Forrás: saját szerkesztés

Az IRF tantárgyból elért metrika értékek, mint eredményváltozó és a SzfI és SzfII, illetve az előfeltétel tárgyakból elért bukásból képzett dummy változó, mint magyarázó változók alapján készített regressziós modell eredménye látható a 7. ábrán. Látható, hogy a SzfI (szfI_model1) változónak nincs szignifikáns hatása az IRF metrika értékeire, azonban a SzfII (szfII_model1) és a bukás változó (elofelt_bukas) a szokásos szignifikancia szinteken szignifikánsak. A koefficiens értékéből leolvasható, hogy a SzfII tárgyból egy egységgel jobb metrika érték elérése esetén az IRF metrika értéke ceteris paribus átlagosan 0,601-del nő, míg az, hogy egy hallgató megbukik előfeltételből, a többi változó változatlanlansága mellett átlagosan 0,597-del csökkenti ennek értékét. Ez tehát azt jelenti, hogy ha két hallgató SzfI és SzfII metrika értéke is megegyezik, de egyikük a jegy megszerzése előtt korábban már megbukott a két tárgy közül legalább az egyikből, akkor neki várhatóan 0,597-del lesz rosszabb a teljesítménye IRF-ből a másik hallgatóhoz képest.

Láthatjuk tehát, hogy a megvizsgált kétféle módszer (előfeltételmatrix, előfeltételes regresszió) alkalmas az előfeltételek jóságának vizsgálatára mind az előkövetelmények összességére nézve, mind pedig egy-egy konkrét tárgyra vonatkoztatva.

7.5. LAD REGRESSZIÓ

A korábbi elemzések eredményei során láthattuk, hogy a meglévő adathalmazt tekintve az egyes tantárgyakból elért eredmények között milyen szoros kapcsolat van a Spearman-féle rangkorrelációs együttható értéke alapján, illetve vizsgáltuk a meglévő előkövetelmény-rendszer jóságát is. Ez utóbbi azonban nehézséget okozott, mivel a vizsgált évfolyamokban csak két kurzus esetén volt előfeltétel tárgy az operatív tantervben. Ez rávilágított arra, hogy nem csak azt kell tudnunk ele-

mezni, hogy egy meglévő tantervi háló jól van-e kialakítva, hanem egy átfogó módszert is ki kellene alakítani, ami a tárgyak közötti tényleges összefüggéseket mutatja meg. A kutatásnak ezen a pontján tehát az volt a célunk, hogy olyan optimális modelleket határozzunk meg, amelyekben az eredményváltozóként megválasztott tárgyról megállapítható, hogy mely más magyarázó változók hatnak rá szignifikáns mértékben. Ezért itt szintén nem leíró statisztikai módszerekkel dolgoztunk, hiszen nem csak a rendelkezésre álló adathalmazról szerettünk volna megállapításokat tenni, hanem arra törekedtünk, hogy általánosságban tudjunk következtetéseket levonni a gazdaságinformatikus képzésről.

7.5.1. A tantárgyak közötti kapcsolat vizsgálatának eszközei

A modell kialakítása során továbbra is elsődleges szempont volt az, hogy a rendelkezésre álló adatok mérési skálája jelentősen korlátozza a használható elemzési módszerek körét. A vizsgálat során a kiindulási pontunk valamilyen regressziós modell volt, azonban mivel a tantárgyak közötti kapcsolatokat a metrikák segítségével szeretnénk volna feltárni, így a klasszikus lineáris megközelítést nem alkalmazhattuk a 4.1.4-es pontban ismertetett okok miatt. Ekkor ismét felmerült a LAD regresszió használatának ötlete, azonban első körben további elemzéseket kellett végeznünk annak érdekében, hogy megállapítsuk, hogy ezt a módszert alkalmazhatjuk-e. A leíró statisztikai elemzések során, amikor a 0-tól 5-ig terjedő skálán vizsgáltuk az érdemjegyeket, nem határoztuk meg a medián értékét, hiszen ilyen kicsi értékkészleten ez nem adna hasznos többletinformációt a vizsgált ismérvekről. Jelen helyzetben azonban már a rendelkezésünkre álltak a metrika értékek, így megvizsgáltuk, hogy az egyes tárgyak milyen értékkészleten mozognak.

25. táblázat: Értékkészlet elemszáma tantárgyanként

Tárgy	Értékkészlet elemszáma
Analízis	59
Informatika	25
Statisztika I.	29
Lineáris algebra	29
Számítástudomány	48
Szoftver-technológia I.	24
Valószínűségyszámítás	36
Adatbázis rendszerek	54
Operációkutatás	41
Statisztika II.	39
Szoftver-technológia II.	18
Informatikai rendszerek fejlesztése	18
Üzleti intelligencia	9

Forrás: saját szerkesztés

A 25. táblázat mutatja, hogy például az Analízis esetén 59 különböző metrika érték jelenik meg a vizsgált három évfolyamban, ezzel szemben az Üzleti intelligencia esetén ez az érték 9. Azonban már ez is elegendő ahhoz, hogy a medián értéke ne biztosan az értékkészlet közepe legyen, így a metrikák segítségével a LAD regresszió alkalmazhatóvá vált.

A LAD regressziós elemzés során meghatározható egy olyan függvény, amely a magyarázó változók segítségével igyekszik a választott eredményváltozót megbecsülni. Számunkra ez azonban még nem megfelelő, hiszen mi egy olyan optimális modellt szerettünk volna meghatározni, amely csak a ténylegesen szignifikáns változókat tartalmazza. Ennek érdekében a backward változoselektációs módszer alkalmazása mellett döntöttük, az illeszkedés jóságát pedig az AIC mutató segítségével mértük (ezeket részletesebben a 4.1.4. fejezetben ismertettük).

A statisztikai módszerek meghatározása után ki kellett választanunk azt az eszközt, amely segítségével az elemzéseket el tudjuk végezni. Korábban már sok esetben használtuk az RStudio-t, így most is ehhez fordultunk, azonban azzal a problémával szembesültünk, hogy nincs olyan csomag, amely tartalmazna olyan eljárást, amire nekünk szükségünk volt. Mivel törekedtünk arra, hogy a legtöbb folyamatot automatizáljuk, így az R-ben meglévő függvények felhasználásával alakítottunk ki egy olyan függvényt (function-t), amely a metrikák alapján meghatározza az optimális modellt. Ehhez a *readexcel*, a *dplyr* és az *L1pack* (Osorio–Wolodzko, 2017) csomagokat használtuk.

8. ábra: LAD regresszió az R-ben

```
LAD_stepAIC <- function(object, subject){
  feltetel <- TRUE
  szamlalo <- 1

  while(feltetel) {
    lad_regression <- lad(paste(subject, "_modell~.", sep=""), data = object)
    lad_sum <- summary(lad_regression)
    aicA <- AIC(lad_regression)
    cat(paste("\n\n", "Step:", szamlalo, "\n", sep=" "))
    print(lad_sum)
    cat(paste("AIC: ", aicA, "\n", sep=" "))

    lad_coeff <- data.frame(lad_sum[["coefficients"]])
    lad_coeff_m <- lad_coeff[-1,]

    szelektalt <- row.names(lad_coeff_m[lad_coeff_m$p.value == max(lad_coeff_m$p.value), ])
    cat(paste("Szelektált változó: ", szelektalt, "\n", sep=" "))

    adatok <- select(object, -ends_with(szelektalt))
    lad_regression2 <- lad(paste(subject, "_modell~.", sep=""), data = adatok)
    lad_sum2 <- summary(lad_regression2)
    aicB <- AIC(lad_regression2)

    if (aicA >= aicB || max(lad_coeff_m$p.value) > 0.15) {
      object <- adatok
      szamlalo <- szamlalo + 1
    } else {
      feltetel <- FALSE
      cat(paste("AIC szelektálás után: ", aicB, "\n", sep=" "))
    }
  }
}
```

Forrás: saját szerkesztés

A LAD_stepAIC function két bemeneti paraméterrel rendelkezik, az object maga az adathalmaz, míg a subject a konkrét tárgyat jelöli. A program alapvetően úgy működik, hogy a kiinduló adatokra meghatározza a LAD regressziós függvényt, és az ehhez tartozó AIC értéket is. Ezt követően kiválasztja azt a magyarázó változót, amelyhez a legmagasabb p-érték tartozik, és ezt ki-szelektálva megismétli az előző lépéseket az új adathalmazon. Az így kapott AIC értéket összeveti a korábbival, és ha ez magasabb, mint volt, akkor ez azt jelenti, hogy megtaláltuk az optimális modellt, amely már nem javítható, a program pedig kilép a ciklusból. Ha pedig alacsonyabb, akkor addig ismétli a lépéseket, amíg el nem jut erre a pontra. A ciklusból való kilépés feltételeként azonban nem csak az AIC értékét vizsgáltuk, hiszen sok esetben azt tapasztaltuk, hogy egy szelekciós lépésnél az Akaike-féle kritérium megnőtt, azonban az optimális modellben maradtak olyan magyarázó változók, melyekhez nagyon magas p-érték társult. A ciklusból való kilépés feltételeként tehát együttesen vizsgáltunk az AIC értékét, és azt, hogy 15%-os p-értékű változónál magasabb ne maradjon a modellben. A változószelekciós algoritmus során azért ezt a magasabb értéket választottuk, mivel 1%-10% p-érték igaz H_0 esetében is előfordulhat (Thompson, 2006). Ezenfelül az esetleges multikollinearitás megnövelheti az érintett együtthatók standard hibáját, így a p-értéket

is. Vagyis a p-érték kapcsán egy szigorúbb kilépési feltétel könnyen azt eredményezhette volna, hogy sok esetben tévesen fogadjuk el H_0 -t. Ez azt jelenti, hogy kisselektálunk olyan változókat is, amelyek valójában nem függetlenek az eredményváltozótól, van köztük kapcsolat. Ez a kapcsolat pedig a szelektálás során nem veszik el, hanem esetlegesen közvetetten, más változókon keresztül jelenik meg, ami hamis képet adhat arról, hogy valójában egy adott magyarázó változó milyen mértékben hat az eredményváltozóra. A 15%-os kritérium megválasztásának az oka tehát az volt, hogy minimalizálni akartuk a másodfajú hiba elkövetésének a valószínűségét, mivel regressziós modellezés során ez a nagyobb értelmezési kockázatot jelentő hibaforrás (Hunyadi–Vita, 2006).

7.5.2. A változók közötti kapcsolatok feltárása

Az elemzéshez használandó statisztikai módszerek és a számítások elvégzéséhez szükséges eszközök már a rendelkezésünkre álltak, így a következő lépésben azt kellett meghatároznunk, hogy pontosan milyen változókat, és hogyan szeretnénk összevetni. Először a tantárgyak közötti összefüggéseket szeretnénk volna feltárni a metrika értékek segítségével, azonban az eredmény- és magyarázó változók meghatározásánál problémába ütköztünk. A rangkorrelációs mátrix kialakításakor egyszerűbb dolgunk volt, hiszen az összes tárgyat viszonyítottuk egymáshoz, most azonban ezt a megközelítést nem tudtuk alkalmazni. A tárgyak egymásra gyakorolt hatását és annak mértékét szeretnénk volna meghatározni, így a vizsgálat szempontjából nem értelmezhető, hogy egy 7. féléves kurzus hogyan hat egy 1. féléves tárgyra, vagyis az időtényezőt, a szemeszterek egymásutániségát is figyelembe kellett vennünk. Ehhez az operatív tanterv nyújtott segítséget, amelyre alapozva fel tudtuk írni, hogy félévről félévre haladva az egyes eredményváltozónak választott tárgyak esetén mely kurzusok jelenjenek meg magyarázó változóként. A 26. táblázat első oszlopa azt mutatja, hogy melyik tárgy melyik félévhez tartozik, ezt támasztja alá a tárgyak színezése is. Az így kapott táblázatot tehát úgy kell értelmezni, hogy például a 2. szemeszterhez tartozó Lineáris algebrához a korábbi féléves tárgyakat, vagyis az Analízist, Informatika alapjait és Statisztika I.-et válogattuk be a kiinduló LAD regressziós modellbe. Itt azonban fontos kiemelni, hogy ez a megközelítés figyelmen kívül hagyja az esetleges endogén hatásokat. Jelen esetben ez azt jelenti, hogy például az Analízis önmagában is és a Lineáris algebrán keresztül is hathat a Valószínűségszámításra. Ezen hatások figyelembevétele további kutatások tárgya lehet.

26. táblázat: Vizsgált változók a LAD regresszióban

		anal	infa	statl	lin	szamt	szfl	valsz	adatb	opk	statll	szfll	irf	ui
1	anal													
1	infa													
1	statl													
2	lin	x	x	x										
2	szamt	x	x	x										
3	szfl	x	x	x	x	x								
3	valsz	x	x	x	x	x								
4	adatb	x	x	x	x	x	x	x						
4	opk	x	x	x	x	x	x	x						
4	statll	x	x	x	x	x	x	x						
4	szfll	x	x	x	x	x	x	x						
5	irf	x	x	x	x	x	x	x	x	x	x	x		
7	ui	x	x	x	x	x	x	x	x	x	x	x	x	

Forrás: saját szerkesztés

Kezdetben a LAD regressziós modell kapcsán kizárólag az egyes tantárgyakhoz tartozó metrika értékek vizsgálatán gondolkoztunk, azonban az adatbázisunkban rendelkezésre álltak még olyan demográfiai adatok, amelyeket úgy gondoltuk, hogy mindenképpen érdemes lenne elemezni. A kiinduló modellek így egészültek ki a nem és az évfolyam változókkal, azonban fontos látni, hogy mind a kettő minőségi ismérv, számmal kódolt kategorikus értékek. Ezeket tehát csak dummy változókként tudtuk kezelni a LAD regressziós elemzések során. A nem esetén a referencia kategória a 0-s (nő) érték lett, és csak plusz egy változóval kellett bővíteni a modelleket. Az évfolyam ennél egy kicsit bonyolultabb volt, hiszen ennél 3 kategória jelent meg, ahogy azt a 27. táblázat mutatja. Ebben az esetben plusz két változót kellett felvennünk: Evfolyam12_13 és Evfolyam13_14, és a 2011/2012-es évfolyam lett a referencia kategória.

27. táblázat: Dummy változó kialakítása az évfolyamokra

	Évfolyam 12_13	Évfolyam 13_14
11-12 (ref.kat)	0	0
12-13	1	0
13-14	0	1

Forrás: saját szerkesztés

A korábbi lépések eredményeként már az összes szükséges eszköz és kiinduló regressziós modell a rendelkezésünkre állt az elemzések elvégzéséhez. A Lineáris algebra példáján keresztül

mutatjuk be, hogy hogyan is működik a megírt változószelekciós algoritmus. Első lépésként RStudio-ban meg kell adni, hogy melyik tárgyat szeretnénk elemezni (az adatbázis kialakítása során alkalmazott rövidítéseket kell használni), majd be kell olvasni Excelből az adatokat. Ezt követően pedig az input paraméterek megadásával lehet meghívni a LAD_stepAIC function-t.

9. ábra: RStudio – a program futtatása

```
tantargy <- "lin"  
adatok <- read_excel("LAD_Regresszió.xlsm", sheet= paste(tantargy , "_adatok", sep=""))  
  
LAD_stepAIC(adatok, tantargy)
```

Forrás: saját szerkesztés

Miután lefutott a kód, a 10. ábrán bemutatott outputot kapjuk. Láthatjuk, hogy első lépésben az összes változó felhasználásával kapunk egy regressziós függvényt, és ennek AIC értéke is megjelenik a táblázatban. Ezt követően a program meghatározza a legnagyobb p-értékkel rendelkező változót, ami a jelen esetben az Evfolyam11_12. Ezt kisselektálva újrafut a ciklusban található kódrész, majd eljutunk a második lépéshez, ahol látszik, hogy már eggyel kevesebb változó jelenik meg. Itt ugyanazt a folyamatot végigjárva az Evfolyam13_14 változó esik ki a modellből. A harmadik lépésben azonban láthatjuk, hogy elérkeztünk az optimális modellhez. Ebben az esetben a Nem attribútumot szelektálná ki az eljárás, azonban ezzel az AIC értéke megnőne, vagyis a modell jósága romlana. Ezenfelül nem maradt olyan változó sem a függvényben, amely 15%-nál magasabb p-értékkel rendelkezik, így, mivel mindkét kilépési kritériumnak megfelel, a program nem fut tovább.

10. ábra: LAD_stepAIC output

Step 1.						
Call:	= paste(subject, "_modell~.", sep = ""), da					
Residuals:						
	Min	1Q	Median	3Q	Max	
	-4.2700	-0.4777	0.0000	0.6229	2.2700	
Coefficients:						
	Estimate	Std.Error	Z value	p-value		
(Intercept)	0.4915	0.2547	1.9293	0.0537		
anal_model1	0.3680	0.0718	5.1281	0.0000		
infa_model1	0.3680	0.0668	5.5072	0.0000		
statI_model1	0.2331	0.0661	3.5233	0.0004		
Nem	-0.2331	0.1411	-1.6520	0.0985		
Évfolyam12_13	-0.1658	0.1772	-0.9361	0.3492		
Évfolyam13_14	-0.2584	0.1726	-1.4972	0.1343		
Degrees of freedom:	480 total;		473 residual			
Scale stimate:	0.9735034					
Log-likelihood:	-633.4654 on		8 degrees		of	freedom
AIC:	1282.9308195965					
Szelektált változó:	Évfolyam12_13					

Step 2.						
Call:	= paste(subject, "_modell~.", sep = ""), da					
Residuals:						
	Min	1Q	Median	3Q	Max	
	-4.0996	-0.5146	0.0000	0.5932	2.2103	
Coefficients:						
	Estimate	Std.Error	Z value	p-value		
(Intercept)	0.3300	0.2112	1.5626	0.1181		
anal_model1	0.3356	0.0719	4.6651	0.0000		
infa_model1	0.3820	0.0657	5.8189	0.0000		
statI_model1	0.2561	0.0662	3.8672	0.0001		
Nem	-0.2047	0.1411	-1.4508	0.1468		
Évfolyam13_14	-0.1253	0.1288	-0.9726	0.3308		
Degrees of freedom:	480 total;		477 residual			
Scale stimate:	0.979674					
Log-likelihood:	-635.0104 on		7 degrees		of	freedom
AIC:	1284.02076492842					
Szelektált változó:	Évfolyam13_14					

Step 3.						
Call:	= paste(subject, "_modell~.", sep = ""), da					
Residuals:						
	Min	1Q	Median	3Q	Max	
	-4.0417	-0.5531	0.0000	0.6210	2.2979	
Coefficients:						
	Estimate	Std.Error	Z value	p-value		
(Intercept)	0.2776	0.2035	1.3645	0.1724		
anal_model1	0.3410	0.0717	4.7555	0.0000		
infa_model1	0.3588	0.0658	5.4498	0.0000		
statI_model1	0.2815	0.0663	4.2460	0.0000		
Nem	-0.2204	0.1415	-1.5576	0.1193		
Degrees of freedom:	480 total;		475 residual			
Scale stimate:	0.979674					
Log-likelihood:	-636.498/on		6 degrees		of	freedom
AIC:	1284.99668065809					
Szelektált változó:	Nem					
AIC szelektálás után:	1287.32158319702					

Forrás: saját szerkesztés

A fenti módszerrel az összes tantárgyra meg tudunk határozni egy optimális modellt, melyek összesített eredményét a 28. táblázat mutatja be. Ennek értelmezését is a Lineáris algebra tárgy segítségével mutatjuk be. A $\beta_0 = 0,28$ mutatja, hogy hol metszi a függvény az y tengelyt, vagyis ez azt jelenti, hogy ha az összes többi változó értéke 0, akkor várhatóan ilyen értékű metrikát szereznek a hallgatók az adott tárgyból. Láthatjuk, hogy a Nem is bekerült a modellbe, azonban 5%-os szignifikancia szinten még nem szignifikáns, így az elsőfajú hiba bekövetkezésének valószínűsége magasabb. Ennek a változónak az együttható értéke azt jelenti, hogy a modellben szereplő három tárgy esetében azonos teljesítményű hallgatókat vizsgálva egy férfi Lineáris algebra metrikája várhatóan -0,22-dal alacsonyabb, mint egy nőé. Ezenfelül láthatjuk, hogy mind a három korábbi féléves tárgy is bekerült a modellbe, így például elmondható, hogy egy 1 metrika értékkel jobb eredmény Analízisből a többi változó változatlansága (a másik két tárgy esetében azonos teljesítményű és nemű hallgató) mellett átlagosan 0,34-dal növeli a Lineáris algebra metrika értékét.

Az eredményeket vizsgálva még fontos kiemelni a Szoftver-technológia I. tárgyat, hiszen itt mind a két évfolyam változó szignifikáns, és ezek együtthatói pozitív előjelűek, és viszonylag magasak is. Ez azt jelenti, hogy két hallgató közül, akik ugyanolyan eredményt értek el Informatikából, Statisztika I.-ből, Lineáris algebrából és Számítástudományból, ha az egyikük a 2013/2014-es évfolyamba járt, akkor átlagosan majdnem egy metrika értékkel jobbat szerzett Szoftver-technológia I.-ből, mint az, aki a 2011/2012-esben kezdett. Azaz megállapítható, hogy a vizsgált három évfolyamban évről évre könnyebbé vált a kurzus.

28. táblázat: LAD regressziós eredmények

	β_0	Nem	Évf. 12-13	Évf. 13-14	anal	infa	statl	lin	szamt	szfl	valsz	adatb	opk	statll	szflil	irf
anal																
infa																
statl																
lin	0.28	-0.22			0.34****	0.36****	0.28****									
szamt	0.00				0.43****	0.18**	0.32****									
szfl	-0.64		0.64**	0.93****		0.45****	0.23*	0.27**	0.31**							
valsz	0.00						0.19**	0.39****	0.32****							
adatb	-0.53					0.16		0.31***		0.21**	0.29**					
opk	-0.11				0.16*		0.12	0.19*	0.20**	0.16**	0.12					
statll	-0.29						0.42****		0.29**	0.18*	0.29**					
szflil	0.78**					0.13			0.123	0.52****	0.18*					
irf	-0.26		0.38**					0.17*				0.21**	0.33****		0.25***	
ui	1.16****											0.18**		0.19***		0.39****

Forrás: saját szerkesztés

A táblázat tehát megfelelően összegzi, hogy mely tárgyak között milyen mértékű kapcsolat van. Fontos azonban kiemelni, hogy habár ezek szignifikáns összefüggések, nem feltétlen beszélhetünk ok-okozati kapcsolatokról. Vagyis biztosan nem tehetünk olyan megállapításokat, hogy ha valaki bizonyos tárgyakból jól teljesített, akkor emiatt fog egy későbbi kurzust is hasonló eredményekkel zárni. Például nem állíthatjuk biztosan, hogy ha valaki Szoftver-technológia II.-ből a legjobb érdemjegyet szerezte, akkor az azért van, mert Szoftver-technológia I.-ből is hasonlóan teljesített. Lehet, hogy ennek hátterében az áll, hogy a hallgató algoritmikus gondolkodásmódja kiváló, ezért nem okoz neki nehézséget a kurzusok teljesítése. Tehát habár ok-okozati összefüggéseket nem tudunk megállapítani, azt biztosan tudhatjuk, hogy mely tárgyak állnak egymással kapcsolatban.

7.5.3. Az IRF és ÜI kurzusok eredményeinek bemutatása

A kapott eredmények gyakorlati értelmezését szintén az Informatikai rendszerek fejlesztése és az Üzleti intelligencia tárgyak segítségével szeretnénk bemutatni. Az IRF az 5. szemeszterben kötelezően teljesítendő kurzus, melynek előfeltétele a Szoftver-technológia II. A kurzus korábbi tárgyak során elsajátított objektumorientált programozási nyelvre építve olyan területekre ad betekintést, mint például a verziókövetés, a fejlesztési minták és az agilis módszertan. Az IRF tehát összefoglalja a korábban tanultakat, és egy átfogó képet ad a szoftverfejlesztés területéről, illetve iránymutatást ad azon hallgatók számára, akik ezen a vonalon kívánnak tovább fejlődni és dolgozni (2. számú melléklet). Ezt támasztják alá a 28. táblázatból leolvasható eredmények is, az IRF-re az alábbi tárgyak hatnak szignifikánsan:

- Lineáris algebra, melynek célja a matematikai készségek fejlesztése, és az alaptudás megszerzése.
- Adatbázis rendszerek, melynek keretein belül az MSSQL nyelv, és az adatbázis ismeretek elsajátítása a cél.
- Operációkutatás, melynek során a lineáris programozás alapjaival ismerkedhetnek meg a hallgatók.
- Szoftver-technológia II., melynek célja az objektumorientált programozási nyelv ismeretének elmélyítése.

Jól látszik tehát, hogy az IRF tényleg egy olyan tárgy, mely nagyban építkezik a korábbi félévekben elsajátított tudásra. Fontos azonban kiemelni, hogy ez egy olyan kurzus, amelyet, ha

sikeresen akarnak teljesíteni a hallgatók, akkor matematikai, adatbázis kezelési és programozói tudásra is szükségük van, vagyis valóban összefoglalja a korábban tanultakat minden területről. A táblázatból továbbá az is kiolvasható, hogy a négy tárgyváltozó közül három 4. féléves kurzus, vagyis közvetlenül az IRF előtt tanulják a hallgatók. Egyik oldalról ez azt jelenti, hogy a képzés struktúrája jól van felépítve, hiszen a tárgyak folyamatosan építkeznek egymásra. Azonban itt sem szabad megfeledkezni a 7.5.2-es pontban említett endogén hatásokról (például a Számítástudomány az Operációkutatáson keresztül hathat az IRF-re), de ezeket ennek a kutatásnak a keretein belül nem tárgyaljuk. Ezenfelül ebben a modellben is szignifikánsan jelenik meg az egyik évfolyamváltozó, vagyis azt is le tudjuk olvasni, hogy 2012/2013-as évfolyamban a hallgatók átlagosan jobb eredményeket értek el. Ez jó visszajelzés lehet az oktató számára, hiszen, ha például történt módosítás a követelményrendszerben, vagy ha új oktató érkezett az egyetemre, akkor lehet mérni, hogy a tárgy teljesíthetőségén ez hogyan változtatott.

Ezekén felül fontos kiemelni az együttthatók értékét is, hiszen ezen adatok is nagyon sok információt tartalmaznak. Jelen esetben az Operációkutatás hat a legnagyobb mértékben az IRF-re, de az Adatbázis rendszerek és a Szoftver-technológia II. együttthatói sem alacsonyak. Ezzel szemben a Lineáris algebra az, amely a legkisebb mértékben (0,17) befolyásolja az eredményváltozót. Ezek az együttthatók azért összehasonlíthatóak, mert minden változó ugyanazon a metrikán mér. Továbbá a β_0 értéke -0,26, ami első ránézésre azt jelentheti, hogy ez egy nehezen teljesíthető tárgy, hiszen az eredményváltozó értékét egy konstans értékkel mindig csökkentjük. Azonban ezt összevetve a többi együtttható értékével, melyek magasak, arra a következtetésre juthatunk, hogy a tárgyat teljesíteni önmagában nehéz, de ha a szignifikáns kurzusokba megfelelő energiát fektetünk, és jobb eredménnyel zárjuk őket, akkor könnyedén szerezhethünk IRF-ből is megfelelő bejegyzést. A leíró statisztikai elemzések kapcsán kapott eredményekhez visszacsatolva, ahol, habár mint sokaságot vizsgáltuk az adathalmazt, szintén arra a megállapításra jutottunk, hogy az Informatikai rendszerek fejlesztése tárgy teljesíthetősége úgy van kialakítva, hogy képes a hallgatók között különbséget tenni. A vizsgálatok eredményeképpen megállapítható, hogy a korábbi alapozó tárgyak során elsajátított tudás nagyban hozzájárul ahhoz, hogy eredményesen el tudjuk végezni az IRF kurzust.

Az Üzleti intelligencia egy szakmai kötelező tárgy, melyet a hallgatók az utolsó félévben teljesítenek. A tantárgyi adatlap szerint a kurzus áttekintést nyújt a legfontosabb üzleti intelligencia

megoldásokról és azok területeiről, melynek során a hallgatók megismerkednek olyan rendszerekkel, mint az OLAP, az üzleti elemzések, a riportok és az adatbányászati megoldások. Az IRF-hez hasonlóan ez a tárgy is a korábbi alapozó kurzusok során megszerzett tudásra épít, azonban ez inkább az adatelemzés és -vizualizáció területén kíván átfogó képet nyújtani (3. számú melléklet).

A 28. táblázatot megvizsgálva láthatjuk, hogy az ÜI-re az Adatbázis rendszerek, a Statisztika II. és az Informatikai rendszerek fejlesztése tantárgyak hatnak szignifikánsan. Tehát ebben az esetben is elmondható, hogy a kurzus három eltérő területre építkezik, melyek a statisztikai, szoftverfejlesztési és adatbázis ismeretek. Az ábrán az is jól látszik, hogy ebben az esetben nem a közvetlenül megelőző féléves tárgyak azok, amelyek hatással vannak az ÜI metrika eredményekre. Azonban megállapítható, hogy ezek a kurzusok már nem alapozó tárgyak, tehát ebben az esetben is feltételezhetjük, hogy a tantárgyak sorrendje jól van kialakítva.

Az együttthatókat vizsgálva azonban az Üzleti intelligenciával kapcsolatban más következtetéseket tudunk levonni, mint az IRF esetében. A β_0 értéke 1,16 vagyis elmondható, hogy ha a szignifikáns tárgyakat még csak fel sem vesszük, abban az esetben is várhatóan teljesíteni tudjuk a tárgyat. Az IRF együttthatójának kivételével a többi együtttható értékén pedig az látszik, hogy a szignifikáns változók nagymértékben nem befolyásolják az ÜI metrika értékének alakulását. Számunkra ez tehát azt jelenti, hogy az Üzleti intelligencia egy teljesíthető tárgy, amely, habár építkezik máskurzusokra, azok nem hatnak rá nagymértékben, egyedül az IRF metrika értéke az, amely jelentősen befolyásolja az eredményváltozót.

Az eredményeket a leíró statisztikai elemzéssel összevetve érdekes megállapításokat tehetünk. A LAD regressziós függvényt vizsgálva azt mondhatjuk, hogy az Üzleti intelligenciát könnyű teljesíteni, azonban a három évfolyamban a jegyek eloszlása közel normális volt. Ez viszont azt jelenti, hogy bár a tárgyból könnyű elégséges érdemjegyet szerezni, a kurzus képes megkülönböztetni a hallgatói teljesítményeket, és jeles bejegyzést már csak kevesen tudnak elérni. Ezenfelül azt is láttuk, hogy habár három tantárgy hat rá szignifikánsan, összességében az ÜI nem épít olyan nagymértékben a korábban megszerzett tudásra, mint az IRF.

7.6. KÖVETKEZTETÉSEK ÉS JAVASLATOK

Az optimális LAD regressziós modellek meghatározása után kidolgoztunk egy előfeltétel-javaslatot a Gazdaságinformatikus BSc képzés tantervi hálójára, amely irányított gráfként reprezentálható. Az gráf csúcsaiban öt szinten, lentől felfelé haladva félévenkénti csoportosításban ábrázoltuk a vizsgált tárgyakat. Az ÜI és az IRF, bár más félévhez tartoznak, egy szintre kerültek, mivel az ÜI a képzés „adatelemzés”, az IRF pedig a „fejlesztés” vonalának „végállomása”. A gráf élei a kurzusok közötti kapcsolatokat jelenítik meg. Kezdetben a 28. táblázat alapján a tantárgyankénti regressziós modellekben szereplő összes hatást megjelenítettük. Ezután eltávolítottuk az 5%-os szignifikancia szinten nem szignifikáns kapcsolatokat, hiszen ezekben az esetekben nagyobb a valószínűsége, hogy egy előkövetelmény-feltétel létrehozását tévesen ítéljük meg.

Következő lépésben tantárgyanként haladva vizsgáltuk meg a gráfban élként ábrázolt kapcsolatokat és háromféleképpen döntöttünk ezekről: 1) az élt eltávolítottuk és nem szerepeltettük az ábrán, ha a magyarázó változó hatása nem jelentős, vagy nem szignifikáns; 2) az élt szaggatott szürke vonallal jelöltük azokban az esetekben, ahol már jelentős a hatás az eredményváltozóra, azonban az előkövetelmény létrehozása szakmailag nem indokolt; 3) fekete nyíllal jelöltük azokat, ahol érdemes elgondolkodni a korábbi féléves tárgy előkövetelményé tételét.

A mérlegelés során első lépésben azt vizsgáltuk, hogy a regressziós modellekben a magyarázó változóként szereplő tárgy milyen szignifikancia szinten szignifikáns. Abban az esetben, ha egy magyarázó változó csak 5%-on mondható szignifikánsnak, a hozzá tartozó élt eltávolítottuk. Erre jó példa a Lin és az IRF közötti kapcsolatot, melyet kivettünk a modellből. Ennél kisebb p-érték esetén, amennyiben nem volt egyértelmű a döntés, más tényezőt is figyelembe kellett vennünk. Ezért megvizsgáltuk minden esetben a regressziós együtthatók értékét, azaz, hogy a magyarázó változó milyen mértékben hat az eredményváltozóra. Ha ez az érték alacsony, tehát a korábbi féléves tárgyból nyújtott teljesítmény csak kis mértékben hat a vizsgált kurzus eredményére, akkor nem érdemes a kettő között előfeltételt létrehozni. Például bár a StatI hatása a Valsz-ra 1%-os p-értéken szignifikáns, azonban a regressziós együttható értéke kicsi, csak 0,19, így a kettő közötti kapcsolatot nem ábrázoltuk. Így tehát a p-érték és a regressziós együttható alapján döntöttük el azt, hogy a tárgyak között ábrázoljunk-e élt a gráfban. Mivel ez a két adat nem minden esetben adott elég információt arra vonatkozóan, hogy a két tárgy között kell-e előfeltételnek lennie, ezért a tényleges előkövetelmény-kapcsolatok kiválasztásához figyelembe vettük a tantárgyak során oktatott ismeretanyagok egymásra épülését is. Ezenfelül elemeztük, hogy a vizsgált adatokon mekkora volt

a két tárgy közötti korrelációs együttható értéke, hogy a három évfolyamon az eredmények közti együttmozgás mértéke összhangban van-e a feltételezett előkövetelmény felállításával. Így tehát a négy figyelembe vett szempont a következő volt:

- a magyarázó változó 1%-on szignifikáns
- a regressziós együttható értéke előfeltételes kapcsolatnál legalább 0,3
- a rangkorrelációs együttható értéke a vizsgált évfolyamokon legalább közepesen erős
- a kurzuson oktatott anyagok épülnek egymásra

Ehhez a 7.3.2. fejezetben bemutatott *Összes null adatbázis c) metrika* értékeire vonatkozó rangkorrelációs mátrixot úgy alakítottuk át, hogy minden tárgy esetében csak azokat a kapcsolatokat jelenítettük meg, amelyek korábbi féléves tárgyakhoz tartoznak. Emellett eltávolítottuk belőle azokat az értékeket, melyek az optimális regressziós modellekbe nem kerültek be, illetve azokat, amik 5%-os szignifikancia szinten nem voltak szignifikánsak. Így jött létre a 29. táblázat.

**29. táblázat: Rangkorrelációs mátrix a regressziós modellek alapján
szignifikáns kapcsolatokkal**

	anal	infa	statl	lin	szamt	szfl	valsz	adatb	opk	statll	szfll	irf
anal												
infa												
statl												
lin	0,5762	0,4845	0,5562									
szamt	0,5923	0,4442	0,5873									
szfl		0,4492	0,4516	0,3470	0,4132							
valsz			0,4998	0,4941	0,4974							
adatb				0,4390		0,3853	0,4282					
opk	0,4985			0,4365	0,5246	0,4338						
statll			0,5477		0,5056	0,4760	0,4863					
szfll					0,3244	0,4730	0,3623					
irf				0,3861				0,5214	0,5047		0,4076	
ui								0,4372		0,5097		0,5207

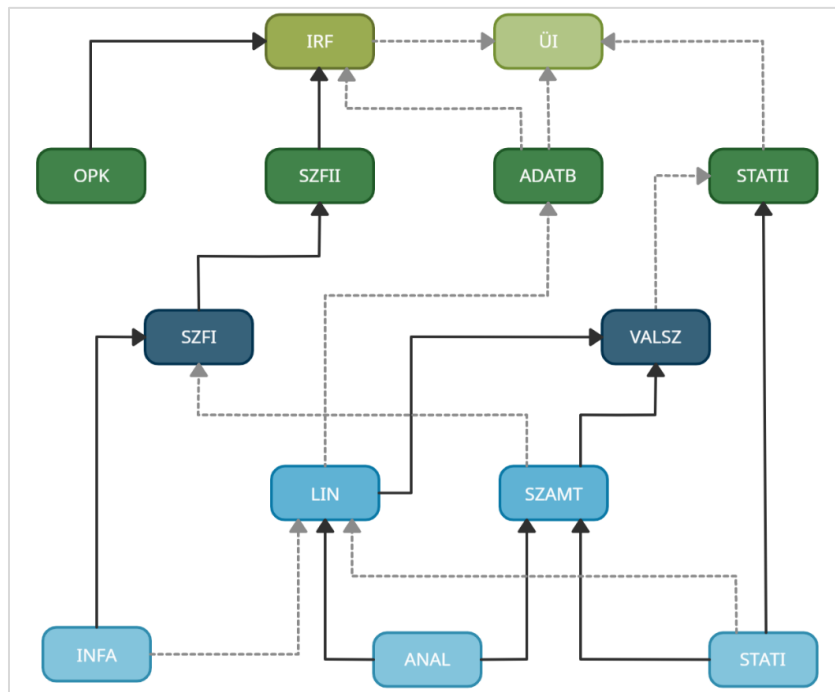
Színmagyarázat: sárga: $0,3 \leq \rho_{on} < 0,5$; világoskék: $0,5 \leq \rho_{on} < 0,7$

Forrás: saját szerkesztés

Ez a négy tényező már elegendő információt nyújtott ahhoz, hogy minden legalább 5%-on szignifikáns kapcsolat esetében eldöntsük, hogy ábrázoljuk-e ezeket a gráfon, és ha igen, akkor milyen módon. Ennek megfelelően az összes tárgyon végighaladtunk, és megvizsgáltuk az összes kapcsolatot annak érdekében, hogy eldöntsük, hogy ezek miként szerepeljenek a tantervi hálót rep-

rezentáló gráfban. A vizsgálat eredménye a 11. ábrán látható. Az Anal-Lin-Valsz-StatII-ÜI kurzusokon keresztül mutatjuk be, hogy pontosan milyen tényezőket és hogyan mérlegeltünk a döntés meghozatalakor.

11. ábra: Előfeltétel-rendszer javaslat



Forrás: saját szerkesztés

Elsőként az Analízis Lineáris algebrára gyakorolt hatását vizsgáltuk. Az Anal és a Lin közötti kapcsolat 0,01%-on szignifikáns, a regressziós együttható értéke pedig 0,34, ami kicsivel több mint egyharmad metrika értéknek felel meg. Ezen két információ alapján a két kurzus között ábrázolt kapcsolatot megtartottuk. A két tantárgyból elért hallgatói eredmények együttmozgása is kiemelkedő volt a vizsgált évfolyamokon, a rangkorreláció együttható mértéke 0,58. Mivel mindkét kurzus matematikai alapozó tárgynak számít, és a vizsgálat alapján jelentős mértékű hatást állapítottunk meg, ezért érdemes lehet elgondolkozni azon, hogy az Analízis teljesítése nélkül a hallgatók ne kezdhessék el a Lineáris algebrát tanulni.

A következő vizsgált tárgy a Valószínűségszámítás. A Lineáris algebra kurzuson nyújtott teljesítmény a regressziós modell alapján 0,1%-on szignifikáns mértékben hat a Valsz eredményekre. A regressziós együttható, azaz a hatás mértéke 0,39, ami szintén jelentősnek mondható.

Mivel mindkét tárgy matematikai alapozó kurzus, ezért építhetnek egymásra. A rangkorreláció értéke a vizsgált adatokon itt valamivel kevesebb, 0,50. Ebben az esetben is jogosan merülhet fel az előfeltétel kapcsolat kialakításának szükségessége, így a gráfban az élt sima fekete nyílként jelöltük.

A Valószínűségszámítás 1%-on hat szignifikáns mértékben a StatII eredményekre, és a regressziós együttható is csak szűk harmad metrika értéknyi, 0,29. A két adat alapján nem dönthető el egyértelműen, hogy a két tárgy közti kapcsolatot ábrázolni kell-e. A StatII kurzuson oktatott anyag azonban épít a valószínűségszámítási alapokra, ezért, ha nem is egy szoros kapcsolat, de gyengébb mindenképpen van. A korrelációs együttható értéke ebben az esetben 0,36, ami szintén nem kiemelkedő. Ezért úgy döntöttünk, hogy konkrét előfeltételként nem ábrázoljuk, de szaggatottal jelöljük a kettő közötti egymásra épülést.

Végül az utolsó tárgy a legutolsó féléves Üzleti intelligencia. A regressziós modelljében a StatII mint magyarázó változó 0,1%-os p-értéken szignifikáns, az együtthatója pedig 0,19. Ez az érték nem jelentős hatás, azonban a két tárgyból elért hallgatói eredmények közötti korrelációs együttható 0,52. Emellett az ÜI során oktatott adatelemzés jelentős mértékben épít a statisztikai ismeretekre, így szaggatott vonallal jelöltük a kettő közötti kapcsolatot. Előfeltételként nem határoznánk meg a Statisztika II.-t azonban mindenképpen szem előtt kell tartani, hogy a két tárgy között van összefüggés.

A nem részletezett tárgyak közötti kapcsolatokat is hasonló alapokon alakítottuk ki, vagy távolítottuk el. Az előfeltétel-javaslat kialakításánál törekedtünk arra, hogy ne hozzunk létre olyan kapcsolatokat, ahol egy kurzus közvetlen és közvetett (másik tárgyon keresztül) előkövetelményként is bekerül egy későbbi félévben oktatott tárgyhöz.

Az előfeltétel-rendszer kialakításánál célszerű figyelni arra, hogy bár viszonylag sok esetben láthatunk magas megbízhatósági szinten is szignifikáns értékeket, de ezek, ahogy már említettük nem jelentenek ok-okozati kapcsolatot. Emellett, az előfeltétel tárgyak számát nem érdemes túlságosan magasra emelni, mert úgy a hallgatók, ha egy tárgyat elsőre nem teljesítenek, akkor akár több félévet is csúszhatnak a képzés elvégzésével.

Összegezve tehát az egyes elemzési pontokon végighaladva meg tudunk határozni a vizsgált évfolyamokra olyan előkövetelmény-rendszer javaslatot, melyben a tárgyak közti valós kapcsolatokat látjuk. A 3. fejezetben bemutatott tanulóelemzési lépéseknél azonban láthattuk, hogy ez egy

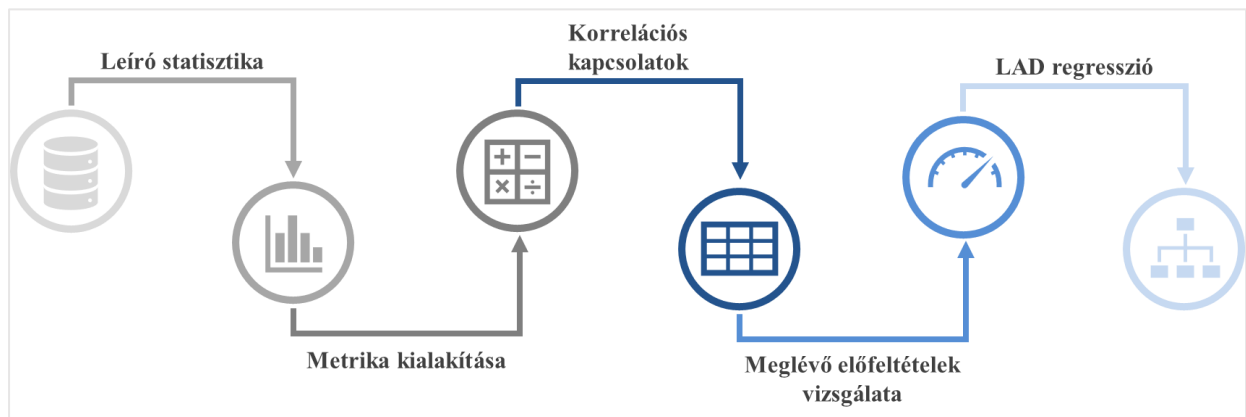
iteratív folyamat. Azaz az elemzések eredményei alapján kialakított előkövetelményeket a képzésbe implementálva az újabb hallgatói eredményekből visszajelzés készíthető. Az újonnan képződött adatokat visszaépítve a folyamatba monitorozható és tovább tökéletesíthető a már kialakított rendszer.

8. ÖSSZEFOGLALÁS

Az osztályzatelemzés a hallgatók érdemjegyei alapján, statisztikai módszerek alkalmazásával próbál információkat feltárni a hallgatók eredményével és a vizsgált képzéssel kapcsolatban. Dolgozatunk célja olyan keretrendszer létrehozása volt, melynek elemei az osztályzatelemzési folyamat alapjait képezhetik. Ehhez a kutatásunk során végighaladtunk a tanulóelemzés azon lépésein, amelyek alkalmasnak bizonyultak arra, hogy a felsőoktatási intézményekben rendelkezésre álló Neptun rendszer, vagy egyéb hagyományos hallgatói adatok tárolására alkalmas program által gyűjtött hallgatói eredményekre vonatkozó adatokból osztályzatelemzési modellt építsünk. A kutatásunk végén megállapíthatjuk, hogy a hipotézisünk beigazolódott, lehetséges olyan, kizárólag Neptun adatokra támaszkodó osztályzatelemzési keretrendszert kialakítani, melynek célja a képzési struktúra javítása.

A megalkotott keretrendszer öt fő elemét a 12. ábrán láthatjuk.

12. ábra: Osztályzatelemzési keretrendszer folyamatábrája



Forrás: saját szerkesztés

A keretrendszer magában foglalja az adatok megismerésére vonatkozó leíró statisztikai eszközöket, melyek az egyes tárgyak során nyújtott hallgatói eredményekről adnak átfogó képet. Ennek segítségével megvizsgálható, hogy melyik évfolyamban mely tárgyak voltak összességében

könnyebben vagy nehezebben teljesíthetők, akár érdemjegy, akár tárgyfelvétel- és vizsgaszám szempontjából. A következő eszköz a teljesítmény-metrika, amelynek használatával egy változóban tárolunk el több adatot a hallgatói teljesítményre vonatkozóan. Így az elért eredményeket egy sokkal szélesebb értékkészleten tudjuk elhelyezni, mint azt az érdemjegyek teszik. Ezzel nagyobb különbségeket alakítottunk ki a hallgatók teljesítménye között, megteremtve a lehetőséget komplexebb elemzésekre. A harmadik eszköz a rangkorreláció, mellyel megvizsgálható a tantárgyakból elért eredmények közötti kapcsolatok szorossága. Megfigyelhető, hogy melyek azok a kurzusok, amelyek között jelentős kapcsolat van, a teljesítmény-metrikának köszönhetően már nem csak az egyes változókra, hanem az összetett eredményekre vonatkozóan is. A negyedik lépésben az előfeltételek hatását vizsgálhatjuk, és ha nincs megfigyelhető kapcsolat az előfeltétel léte és a metrika értékek között, akkor érdemes lehet felülvizsgálni, hogy valóban szükséges-e az az előkövetelmény. A keretrendszer utolsó elemeként a tárgyakból elért érdemjegyek tényleges, közvetlen egymásra hatását figyelhetjük meg backward változószelekciós algoritmussal támogatott LAD regresszió segítségével. Ezzel kiépíthető egy előkövetelmény-rendszer, vagy ellenőrizhető a már meglévő előfeltételek jósága.

Az így kidolgozott keretrendszer elemei lehetővé teszik, hogy az elemzéseket a hagyományos tanulásirányítási rendszerekben rendelkezésre álló adatokon végezzük el. Az eszközkészlet meghatározásakor figyelembe vettük, hogy a hallgatói érdemjegyeket sorrendi skálán értelmezzük, és emiatt a statisztikai eszközök speciális változatait kell használni. Emellett a hallgatói teljesítményt nem csak az érdemjegyek, hanem az azokhoz elvezető út alapján is értékelik a keretrendszerünk elemei, ezzel hangsúlyozva, hogy értékesebb az a jegy, amit kevesebb próbálkozással sikerül megszerezni. Az osztályzatelemzési folyamat végén előálló LAD regressziós modellek a backward változószelekciós algoritmusnak köszönhetően optimális modellek lesznek, hiszen csak a releváns (szignifikáns) változókat fogják tartalmazni. Ezek a regressziós modellek segítenek abban, hogy az előkövetelmény-rendszert ne csak az alapján alakítsuk ki, ahogy a tananyag szerint megfelelőnek gondoljuk, hanem vegyük figyelembe az első látásra talán nem egyértelmű, de elemzésekkel alátámasztható összefüggéseket is.

A kutatásunk eredményeképpen előálló keretrendszer az osztályzatelemzési folyamatot támogatja. Ezzel a felsőoktatási intézmények szakfelelősei, szakfejlesztői feltárhatják a tantárgyak közötti tényleges kapcsolatokat, és valós egymásra épülésen alapuló előkövetelmény-rendszert hozhatnak létre. Az új előfeltételek javaslata mellett a már meglévő előkövetelmények jóságát is

segítik felmérni az előfeltételvizsgálatok és az optimális regressziós modell. Ezzel elkerülhető, hogy felesleges előkövetelményekkel gátolják a hallgatók előrehaladását, azonban ne engedjék továbblépni őket, amíg a szükséges készségeket legalább elégséges szinten el nem sajátították. Így közvetett módon az oktatók is érdekeltek a kapott eredményekben, hiszen ezáltal nem kerülnek hozzájuk olyan hallgatók, akik a megelőző kurzusokon nem szerezték meg a kellő tudást. A tárgyak közötti kapcsolatfeltárás mellett a leíró elemzések alkalmasak arra is, hogy egy-egy tárgyról visszajelzést adjanak, és ha a kapott eredmények alapján látszik, hogy például a teljesíthetőség nem megfelelő, akkor ez időben cselekvésre indíthatja a döntéshozókat. Emellett a LAD regressziós módszertan alkalmas arra, hogy rávilágítson, ha a vizsgált évfolyamok eredményei között szignifikáns eltérések vannak. Ezzel érzékelhetővé válik, ha változott a követelményrendszer, vagy új oktató folytatta a tárgy tanítását. A visszajelzéseknek köszönhetően akár pozitív, akár negatív irányba is változtatni lehet a feltételeken.

Az így kidolgozott keretrendszert a felsőoktatási intézmények szakfelelősei, szakfejlesztői, tárgyfelelősei és oktatói használhatják a képzések javítására és a képzési struktúra fejlesztésére. Jelen kutatásban a Gazdaságinformatikus BSc képzést vizsgáltuk, és az összeállított módszertan segítségével kidolgoztunk egy előfeltétel-rendszer javaslatot. A keretrendszer azonban megfelelően alkalmazható minden felsőoktatási intézetben, hiszen ahogyan arra már többször kitértünk, a kutatás alapvető célja az volt, hogy kizárólag az adminisztrációs rendszerekben rendelkezésre álló hallgatói adatok alapján tudjunk elemzéseket végezni és javaslatokat tenni. Ennek oka, hogy sok felsőoktatási rendszerben nincsen olyan tanulóelemzési platform, amelyből kinyerhetőek lennének a komplex tanulóelemzési folyamatokhoz szükséges adatok. Előfordulhat az is, hogy bár ezek a felületek léteznek, de nincs lehetőség vagy kapacitás arra, hogy összetett folyamatokon keresztül elemezzék az adatokat. Ez a módszer viszonylag kevés eszköz felhasználásával már tud olyan eredményeket adni, amelyek által képzést fejlesztő döntések hozhatók.

Az osztályzatelemzési keretrendszer és folyamat több lehetséges kutatási irányba is elágaztatható. A folyamat könnyebbé tételéhez és gyorsításához érdemes lehet hosszabb távon automatizálni azt, oly mértékben, hogy az értelmezéseken kívül a lehető legkevesebb emberi beavatkozással lehessen megoldani az egyes lépéseket az adatok begyűjtésétől kezdve az adattisztításon át az elemzésekig. Ez főként akkor kiemelt fontosságú, ha az osztályzatelemzési vizsgálatokat rendszeresen, akár évről évre el akarjuk végezni a képzés monitorozásának céljából. További kutatási irány a vizsgálat elmélyítésére, hogy milyen egyéb változókkal lehet bővíteni az itt bemutatott regressziós

modelleket. Többek között vizsgálható, hogy milyen hatása volt esetlegesen az oktató vagy a tanterv változásának. Végül újabb vizsgálat tárgyát képezheti az endogén hatások elemzése. Vagyis az, hogy a regressziós modellben jelzett hatások milyen mértékben függenek a többi változótól, hogyan befolyásolják egymást a magyarázó változók és milyen mértékű a multikollinearitás.

IRODALOMJEGYZÉK

Szakirodalom

Baradwaj, B. K. - Pal, S. (2011): Mining Educational Data to Analyze Students Performance.

International Journal of Advanced Computer Science and Applications. Vol. 2. No. 6, pp. 63-69.

URL: <https://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf>

Letöltés dátuma: 2021. március 18.

Charitopoulos, A. - Rangoussi, M. - Koulouriotis, D. (2020): On the Use of Soft Computing

Methods in Educational Data Mining and Learning Analytics Research: a Review of Years 2010–2018. International Journal of Artificial Intelligence in Education, Vol. 30. No. 3, pp. 371-430.

URL: <https://link.springer.com/article/10.1007%2Fs40593-020-00200-8>

Letöltés dátuma: 2021. február 13.

Dyckhoff, A. L. - Zielke, D. - Bültmann, M. - Chatti, M. A. - Schroeder, U. (2012): Design and Implementation of a Learning Analytics Toolkit for Teachers. Educational Technology & Society. Vol. 15. No. 3, pp. 58-76.

URL: <http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=7&sid=73cfcb41-060e-4e83-b4d0-20df7031ea40%40sessionmgr4008>

Letöltés dátuma: 2021. március 22.

Dziuban, C. - Moskal, P. - Cavanagh, T. - Watts, A. (2012): Analytics that Inform the University:

Using Data You Already Have. Journal of Asynchronous Learning Networks, Vol. 16. No. 3, pp. 21-38.

URL: <https://files.eric.ed.gov/fulltext/EJ982670.pdf>

Letöltés dátuma: 2021. február 2.

Elias, T. (2011): Learning Analytics: Definitions, Processes and Potential

URL: <https://landing.athabasca.ca/file/download/43713>

Letöltve: 2021. március 22.

- Fajszai Bulcsú - Cser László - Fehér Tamás (2010): Üzleti haszon az adatok mélyén. Alinea Kiadó, Budapest
- Ferguson, R. (2012): Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, Vol. 50. No. 6, pp. 304–317.
URL: https://oro.open.ac.uk/36374/1/IJTEL40501_Ferguson%2520Jan%25202013.pdf
Letöltés dátuma: 2021. február 13.
- Fitzenberger, B. - Koenker, R. - Machado, J. A. (eds., 2013): Economic Application of quantile regression. Springer Science & Business Media.
- Florence, M. - Abdou, N. (2016): Using Learning Analytics to Assess Student Learning in Online Courses. Journal of University Teaching & Learning Practice. Vol. 13. No. 3, pp. 1-20.
URL: <https://files.eric.ed.gov/fulltext/EJ1110545.pdf>
Letöltés dátuma: 2021. március 22.
- Foster, L. - Diamond, I. - Banton, J. (2014): Beginning statistics: An introduction for social scientists. Sage.
URL: <http://methods.sagepub.com/book/beginning-statistics>
Letöltés dátuma: 2021. február 20.
- Hauke, J. - Kossowski, T. (2011): Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Questiones geographicae, Vol. 30. No. 2, pp. 87-93.
URL: https://www.researchgate.net/publication/227640806_Comparison_of_Values_of_Pearson's_and_Spearman's_Correlation_Coefficients_on_the_Same_Sets_of_Data
Letöltés dátuma: 2021. február 19.
- Hicks, A. - Graber, A. (2010): Shifting paradigms: teaching, learning and Web 2.0. Reference Services Review, Vol. 38. No. 4, pp. 621-633.
URL: <https://www.emerald.com/insight/content/doi/10.1108/00907321011090764/full/pdf?title=shifting-paradigms-teaching-learning-and-web-20>
Letöltés dátuma: 2021. március 22.
- Hunyadi László - Vita László (2006): Statisztika közgazdászoknak. harmadik ed. KSH, Budapest
- Hunyadi László - Vita László (2008): Statisztika II. Aula Kiadó, Budapest

- McNaughton, M. - Rao, L. - Mansingh, G. (2017): An agile approach for academic analytics: a case study. Journal of Enterprise Information Management. Vol. 30. No. 5, pp. 701-722.
URL: <https://www.emerald.com/insight/content/doi/10.1108/JEIM-06-2016-0121/full/pdf?title=an-agile-approach-for-academic-analytics-a-case-study>
Letöltés dátuma: 2021. március 18.
- Mining, T. E. D. (2012): Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.
URL: <https://mobile.eduq.info/xmlui/bitstream/handle/11515/35829/enhancing-teaching-learning-through-educational-data-mining-learning-analytics-us-department-education-2018.pdf?sequence=2&isAllowed=y>
Letöltve: 2021. március 21.
- Saqr, M (2018): Using Learning Analytics to Understand and Support Collaborative Learning
URL: https://www.researchgate.net/publication/336374701_Using_Learning_Analytics_to_Understand_and_Support_Collaborative_Learning
Letöltve: 2021. február 20.
- Scapin, R. (2015): Learning Analytics in Education: Using Student's Big Data to Improve Teaching
URL: https://www.reptiq.ca/wp-content/uploads/2014/10/2015-04_Learning_Analytics.pdf
Letöltve: 2021. február 19.
- Shum, S. B. - Ferguson, R. (2012): Social Learning Analytics. Educational Technology & Society, Vol. 15. No. 3, pp. 3–26.
URL: <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=195c4a74-2e4f-409d-89eb-8985d99890ee%40sessionmgr102>
Letöltés dátuma: 2021. február 11.
- Shum, S. B. - Luckin, R. (2019): Learning analytics and AI: Politics, pedagogy and practices. British Journal of Educational Technology. Vol. 50. No. 6, pp. 2785–2793.
URL: <https://bera-journals.onlinelibrary.wiley.com/doi/pdfdirect/10.1111/bjet.12880>
Letöltés dátuma: 2021. március 22.

Thompson, B. (2006): Critique of p-Values. International Statistical Review, Vol. 74. No. 1, pp. 1-14.

URL: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1751-5823.2006.tb00157.x>

Letöltés dátuma: 2021. április 9.

Tsai, Y. S - Poquet, O. - Gašević, D. - Dawson, S - Pardo A. (2019): Complexity leadership in learning analytics: Drivers, challenges and opportunities. British Journal of Educational Technology. Vol. 50. No. 6, pp. 2839–2854

URL: <http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=24291ecc-37a8-4049-929f-e9918a185aac%40sessionmgr4006>

Letöltés dátuma: 2021. március 22.

Venables, W. N. - Ripley, B. D. (2002): Modern Applied Statistics with S. Fourth edition. Springer, New York, NY, USA

URL:

https://www.researchgate.net/publication/224817420_Modern_Applied_Statistics_With_S

Letöltés dátuma: 2021. március 10.

Zar, J. H. (2005): Spearman rank correlation. In: Armitage, P - Colton, T. (eds.): Encyclopedia of Biostatistics. John Wiley & Sons Ltd., Hoboken, NJ, USA

URL: https://www.researchgate.net/publication/227998354_Spearman_Rank_Correlation

Letöltés dátuma: 2021. február 19.

Zheng, G. - Peltsverger, S. (2015): Web Analytics Overview

URL: <https://www.igi-global.com/viewtitlesample.aspx?id=112470&ptid=76156&t=web+analytics+overview>

Letöltve: 2021. március 20.

Zilvinskis, J. - Willis, J. E. (2019): Learning Analytics in Higher Education: A Reflection. A Journal of Scholarly Teaching, Vol. 14., pp. 43-50.

URL: <https://files.eric.ed.gov/fulltext/EJ1222846.pdf>

Letöltés dátuma: 2021. február 11.

Források

Dr. Vas Réka Franciska (2021): Üzleti intelligencia - 2. előadás. Információrendszerek Tanszék. Budapesti Corvinus Egyetem

Egyéb dokumentumok

Osorio, F. - Wolodzko, T. (2017): Routines for L1 estimation. R package version 0.38.19

URL: <http://l1pack.mat.utfsm.cl>

R Core Team (2019): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

URL <https://www.R-project.org/>

RStudio Team (2020): RStudio: Integrated Development Environment for R, Boston, MA.

URL: <http://www.rstudio.com/>

Wickham, H. - Bryan, J. (2019): readxl: Read Excel Files. R package version 1.3.1.

URL: <https://CRAN.R-project.org/package=readxl>

Wickham, H. - François, R. - Henry, L. - Müller, K. (2020): dplyr: A Grammar of Data Manipulation. R package version 0.8.5.

URL: <https://CRAN.R-project.org/package=dplyr>

Jogszabályok

Corvinus (2020): SZMSZ III. KÖTETE HALLGATÓI KÖVETELMÉNYRENDSZER -III.3.

Tanulmányi és Vizsgaszabályzat.

URL: https://www.uni-corvinus.hu/fooldal/egyetemunkrol/szabalyzatok-utasitasok/szabalyzatok/szervezeti-es-mukodesi-szabalyzat/#hallgatoi_jogorvoslati_rendrol_sz_l_szabalyzat_archivum

Letöltés dátuma: 2021. március 12.

MELLÉKLETEK

1. MELLÉKLET: VIZSGÁLT TANTÁRGYAK ÉS RÖVIDÍTÉSEIK

Tantárgy	Rövidítés
Analízis	Anal
Informatika alapjai	Infa
Statisztika I.	StatI
Lineáris algebra	Lin
Számítástudomány	Szamt
Szoftver-technológia I.	Szfl
Valószínűességszámítás	Valsz
Adatbázis rendszerek	Adatb
Operációkutatás	Opk
Statisztika II.	StatII
Szoftver-technológia II.	SzflII
Informatikai rendszerek fejlesztése	IRF
Üzleti intelligencia	ÜI

(Forrás: saját szerkesztés)

2. MELLÉKLET: INFORMATIKAI RENDSZEREK FEJLESZTÉSE TANTÁRGYI ADATLAP A 2014/2015 TANÉVBŐL

(Forrás: Budapesti Corvinus Egyetem archívuma)

Informatikai rendszerek fejlesztése

A tantárgy kódja: 2SZ31NAK13B

A tantárgy megnevezése (magyarul): Informatikai rendszerek fejlesztése

A tantárgy neve (angolul): Systems Analysis and Design

A tanóra száma (Előadás + szeminárium + gyakorlat + egyéb): 2+2 (előadás+gyakorlat)

Kreditérték: 5

A tantárgy meghirdetésének gyakorisága: őszi félév

Az oktatás nyelve: magyar

Előtanulmányi kötelezettségek: Szoftvertechnológia I. és II.

A tantárgy típusa: kötelező

Tantárgyfelelős tanszék: Számítástudományi Tanszék

A tantárgyfelelős neve: Dr. Csetényi Artur

A tantárgy szakmai tartalma: A tantárgy keretében a hallgatók megismerik a szoftverrendszerek fejlesztésének: specifikációjának, tervezésének, implementálásának és menedzselésének legfontosabb kérdéseit, eszközeit és technikáit. A félév során megtanulják, hogyan lehet nagy komplexitású rendszereket költséghatékony módon fejleszteni, amelyek egyúttal eleget tesznek az

üzembiztonság összetett kritériumának. Részletes tárgyalásra kerülnek a szoftverprojekt-menedzsment és a kockázatkezelés technikái. A kurzus egyik legfontosabb célja a rendszerkövetelmények (a rendszer által nyújtott szolgáltatások magas szintű, absztrakt megfogalmazásai) tervezési folyamatának megismerése. A szemináriumokon korszerű, integrált fejlesztőkörnyezetek használatával a hallgatók betekintést nyernek a gyakorlati szoftverfejlesztés különböző részterületeibe, a korábbi tanulmányok kiterjesztéseként egyéni mellett csoportos tervezési - fejlesztési - tesztelési folyamatokba. Az objektumorientált programozási modellre építve a szemináriumi képzés hangsúlyos elemei az adatvezérelt alkalmazások és a többretegű alkalmazásfejlesztés modern platformjai.

Évközi tanulmányi követelmények: A félév folyamán a hallgatók két elméleti zárthelyi dolgozatot írnak, a szemináriumokon pedig feladatokat oldanak meg. A gyakorlatokon való részvétel kötelező!

Vizsgakövetelmény: gyakorlati jegy

Az értékelés módszere: A félév folyamán a hallgatók két elméleti zárthelyi dolgozatot írnak 20-20 pontért, s maximum 20 pontot szerezhetnek a szemináriumon feladatok megoldásával. Az osztályzat a félévközi folyamatos számonkérésen alapul. A félév sikeres zárásának feltétele, hogy mindkét elméleti zárthelyi dolgozat eredménye külön-külön legalább 8 pont legyen (40%).

Az osztályzat az alábbi ponthatár-táblázat szerint kerül kiszámításra:

Ponthatárok	érdemjegy
0 - 29	1 elégtelen
30 - 37	2 elégséges
38 - 45	3 közepes
46 - 52	4 jó
53 - 60	5 jeles

A félév utolsó elméleti óráján lehetőség van hallgatói prezentációk tartására, amelyekkel extra pontokat lehet szerezni.

Elmulasztott elméleti zárthelyi pótlására igazolt távollét esetén, a vizsgaidőszak elején, egyetlen alkalommal van lehetőség. A szemináriumi pontokat a szorgalmi időszakban lehet megszerezni, ezekutólag nem pótolhatók, nem javíthatók. A tantárgy sikeres teljesítésének feltétele, hogy a félévközi számonkérésekből számított összpontszám elérje az elégséges szintet (30 pont), illetve mindkét elméleti zárthelyi dolgozat eredménye külön-külön legalább 8 pont (40%) legyen. Ha valamelyik feltétel nem teljesül, összevont elméleti javító zárthelyi írható a vizsgaidőszak elején, egy alkalommal, 40 pontért. Ekkor az érdemjegy a javító zárthelyi eredménye és a szemináriumi pontok összegéből alakul ki, a fenti ponthatár-táblázat szerint. Az elméleti ZH-k írásakor segéd-eszközök nem használhatók.

Tananyag leírása: Témakörök:

1. Bevezetés. Szoftverfolyamat modelljei I. (vízesés modell, evolúciós fejlesztés, evolúciós és eldobható szoftverprototípusok készítése).
2. Szoftverfolyamat modelljei II.(CBSE, inkrementális fejlesztés, spirális fejlesztés).

3. Tervvezérelt vs. agilis SW fejlesztési módszerek
4. Szoftvertervezés és rendszertervezés kapcsolata. CASE. Szakmai és etikai felelősség.
5. Informatikai rendszerek típusai és alkalmazásaik.
6. Szociotechnikai rendszerek. Szervezetek és számítógéprendszerek. Ősrendszerek.
7. Kritikus rendszerek. Üzembiztonság, megbízhatóság, biztonságosság, védettség.
8. Projektmenedzsment, projektek tervezése és ütemezése. Kockázatkezelés, kockázat azonosítása, elemzése, tervezése.
9. Funkcionális és nem funkcionális szoftverkövetelmények, felhasználói és rendszerkövetelmények.
10. Követelmények tervezésének folyamatai (megvalósíthatósági tanulmányok, nézőpont-orientált feltárás, forgatókönyvek, használati esetek, etnográfia).
11. Rendszermodellek (környezeti modellek, viselkedési, adatfolyam, állapotátmenet, objektummodellek).

Órarendi beosztás: A NEPTUN Hallgatói Információs Rendszer szerint.

Kompetencia leírása: A hallgató a kurzus során elsajátítja azokat az ismereteket, amelyek alkalmasak teszik őt arra, hogy aktív résztvevője legyen közepes- és nagyméretű informatikai rendszerfejlesztési projekteknek, s azokban különböző szerepköröket legyen képes betölteni. A tárgy elméleti része elsősorban nem a programozással kapcsolatos kérdésekre fókuszál, hanem a rendszerelemzés, modellezés és a szoftverkövetelmény-tervezés folyamatára.

Félévközi ellenőrzések: A félév folyamán a hallgatók két elméleti zárthelyi dolgozatot írnak 20-20 pontért. Az első elméleti ZH időpontja a 7. héten, a másodiké a 13. héten lesz. (Az ütemezéstől kényszerű eltérések elképzelhetők). A szemináriumokon a számonkérés folyamatos, a hallgatók aktív részvétele szükséges és elengedhetetlen. A félév sikeres zárásának feltétele, hogy mindkét elméleti zárthelyi dolgozat eredménye külön-külön legalább 8 pont legyen (40%). A zárthelyi dolgozatok külön-külön nem pótolhatók.

Azon hallgatók, akik a félév során nem érik el a minimum követelményt (30 pont), összevont ZH-t írnak a vizsgaidőszakban az elméleti ismeretekből.

A hallgató egyéni munkával megoldandó feladatai: A rendszerelemzés praktikus fogásait komoly gyakorlással kell és lehet elsajátítani. A szemináriumokon feladatmegoldás folyik, ezek befejezése általában önállóan, egyéni munkával történik.

Szak neve: Gazdaságinformatikus (BSc)

Irodalomjegyzék:

Kötelező irodalom:

Sommerville: Szoftverrendszerek fejlesztése, Panem Kiadó, 2002.

Csetényi Arthur: Informatikai rendszerek fejlesztése, In: Informatika alapjai (Csala, Csetényi, Tarlós), pp. 271-436, Computerbooks, 2001.

Ajánlott irodalom:

A tantárgy oktatói: Dr. Csetényi Artur, Németh Zoltán, Mohácsi László

3. MELLÉKLET: ÜZLETI INTELLIGENCIA TANTÁRGYI ADATLAP A 2014/2015 TANÉVBŐL
(Forrás: Budapesti Corvinus Egyetem archívuma)

Üzleti intelligencia

A tantárgy kódja: 2IR32NDK04B

A tantárgy megnevezése (magyarul): Üzleti intelligencia

A tantárgy neve (angolul): Business Intelligence

A tanóra száma (Előadás + szeminárium + gyakorlat + egyéb): 2+2

Kreditérték: 5

A tantárgy meghirdetésének gyakorisága: őszi szemeszterben

Az oktatás nyelve: magyar

Előtanulmányi kötelezettségek: nincs

A tantárgy típusa: kötelező

Tantárgyfelelős tanszék: Információrendszerek Tanszék

A tantárgyfelelős neve: Kő Andrea

A tantárgy szakmai tartalma: A tantárgy célja, hogy áttekintést nyújtson az üzleti intelligencia megoldások elméleti háttéréről, alkalmazási lehetőségeiről, a gyakorlat során fellépő kihívásokról. A tárgy keretében foglalkozunk az adattárházakkal kapcsolatos tudnivalókkal, az adatminőséghez kötődő kérdésekkel és a „big data” kihívásokkal. Megvizsgáljuk az adattárházakra épülő üzleti intelligencia megoldások sajátosságait, alkalmazásuk korlátait, feltételeit. Áttekintjük az adatbányászat, webbányászat és a szövegbányászat alapjait, támogató megoldásait.

Évközi tanulmányi követelmények: A gyakorlati jegy az alábbi feladatokra kapott részpontoszámok összegéből adódik:

- gyakorlati feladatok 50 pont
- elméleti beszámoló 50 pont (min. 25 pontot el kell érni az elégtelentől különböző jegyhez)
- Plusz pontokat (10 pont) lehet szerezni csoportmunkában (2-3 fő) előzetesen egyeztetett téma feldolgozásával és prezentálásával.

A tárgy értékelése félévközi és vizsgaidőszakbeli számonkérés kombinációjaként történik meg.

Vizsgakövetelmény: gyakorlati jegy

Az értékelés módszere: A tárgy gyakorlati jeggyel zárul. A gyakorlati jegy az alábbi feladatokra kapott részpontoszámok összegéből adódik:

- gyakorlati feladatok 50 pont
- elméleti beszámoló 50 pont (min. 25 pontot el kell érni az elégtelentől különböző jegyhez)

Plusz pontokat (10 pont) lehet szerezni csoportmunkában (2-3 fő) előzetesen kiadott feladat megoldásával; illetve téma feldolgozásával és prezentálásával. A gyakorlatokon való részvétel és folyamatos munka a tárgy teljesítésének feltétele. Hiányozni a szemináriumok 25%-ról lehet. A szemináriumi feladatok más módon nem teljesíthetőek (TVSZ 21§ / 5 pont). Ez jelen kurzus esetében 3 hiányzást jelent. Nagyobb mértékű hiányzás esetében a hallgató elégtelen osztályzatot

szerez, melyet javító vizsgával módosíthat. A javító vizsgán a tanszék biztosítja, hogy a közepes szint a javító vizsgán elérhető legyen, azaz az elérhető legjobb eredmény a közepes (70% értékben).

Értékelés (%-ban):

Elégtelen: 0 - 47

Elégséges: 48 - 60

Közepes: 61 - 73

Jó: 74 - 86

Jeles: 87 – 100

Tananyag leírása:

Bevezetés: az üzleti intelligencia alapfogalmai, üzleti intelligencia és döntéstámogatás

Adatgyűjtés problémái, adatminőség

Adattárházak (adatpiacok) jellegzetességei, kialakításuk folyamata

Adattárházak a gyakorlatban

Üzleti analitika - OLAP

Üzleti teljesítménymenedzsment megoldások

Adatbányászat áttekintés

Adatbányászat - R

Webbányászat és szövegbányászat

Az adatbányászat alkalmazása

BI trendek

Összefoglalás/Hallgatói prezentációk I

Összefoglalás/Hallgatói prezentációk II

Órarendi beosztás: 2+2

Kompetencia leírása: A tantárgy keretében a hallgatók áttekintést kapnak a legfontosabb üzleti intelligencia megoldásokról és területekről. A hallgatók felhasználói szinten képesek lesznek üzleti intelligencia megoldásokkal dolgozni, megismerkednek olyan specializált rendszerekkel, mint az OLAP, üzleti elemzések, riportok, adatbányászati megoldások. A hallgatók megismerik, hogy hogyan támogathatja az üzleti döntéshozatalt egy-egy kiválasztott probléma esetében az üzleti intelligencia megoldás és, hogy melyek a főbb alkalmazási területek.

Félévközi ellenőrzések: A félév során a gyakorlatokon egyéni és csoportos feladatmegoldás is történik, mely során az elérhető pontszám 50%-a szerezhető meg. A csoportos feladatmegoldásra csoportos pontszám kerül meghatározásra. A gyakorlaton három alkalommal az elméleti anyagból beszámoló lesz.

A hallgató egyéni munkával megoldandó feladatai:

A félév során egyénileg és csoportosan megoldandó feladatok elkészítése.

Szak neve: Gazdaságinformatikus szak/ Üzleti informatika szakirány

Irodalomjegyzék:

Kötelező irodalom:

1. Kő Andrea: Üzleti intelligencia; in: Döntéstámogató rendszerek 4. fejezet, szerk.: Sántáné-Tóth Edit, Panem Kiadó, 130-144 oldal
2. Bíró Miklós, Sántáné-Tóth Edit: Döntéelméleti alapfogalmak és Kő Andrea, Lovrics László, Sántáné-Tóth Edit: Döntéshozatal és döntéstámogatás; in: Döntéstámogató rendszerek 2. fejezet, szerk: Sántáné-Tóth Edit, Panem Kiadó, 2008) 22-33 oldal, 42-48 oldal, 51-81 oldal és az előadás kiadott irodalom

Ajánlott irodalom:

1. Fajszki Bulcsú – Cser László - Fehér Tamás: Üzleti haszon az adatok mélyén, Alinea Kiadó, 2010; 45 - 89 oldal
2. Inmon, W. H., “Building the Data Warehouse” New York: Wiley, Third ed. 2002 3. Kimball, R.: “Rating Your Dimensional Data Warehouse” www.intelligententerprise.com – 2000
4. Ramesh Sharda, Dursun Delen, Efraim Turban Business Intelligence and Analytics: Systems for Decision Support, 10/E, Prentice Hall 2014

4. MELLÉKLET: GAZDASÁGINFORMATIKUS (BSC) ALAPKÉPZÉSI SZAK 2012/2013. OPERATÍV TANTERVE
(Forrás: Budapesti Corvinus Egyetem archívuma)

Gazdaságinformatikus (BSc) alapképzési szak 2012/2013. operatív tanterve																								
Kód	Tárgynév	Jelleg	Számonkérés	I. évfolyam (2012. szeptemberben kezdett)						II. évfolyam (2011. szeptemberben kezdett)						III. évfolyam						IV.		
				1		Kredit	2		Kredit	3		Kredit	4		Kredit	5		Kredit	6		Kredit	7		Kredit
				ea	sz		ea	sz		ea	sz		ea	sz		ea	sz		ea	sz		ea	sz	
Alapozó kötelező tárgyak						25			20			25			25			23			30			
4MA12NAK12B	Analízis	K	v	2	2	5																		
2SZ31NAK03B	Informatika	K	v	2	2	5																		
4MI25NBK04B	Közgazdaságtan	K	v	2	2	5																		
4ST14NAK04B	Statisztika I.	K	v	2	2	5																		
2VL60NBK01B	Vállalatgazdaságtan	K	v	2	2	5																		
2VE81NGK14B	Vezetés és szervezés	K	v				2	2	5															
4MA12NAK31B	Lineáris algebra	K	v				2	2	5															
2SZ31NAK04B	Számítógép-architektúrák	K	v				2	2	5															
2SZ31NAK12B	Számítástudomány	K	v				2	2	5															
2IR32NAK05B	Információmenedzsment	K	gy							2	2	5												
2SZ31NAK08B	Operációs rendszerek	K	v							2	2	5												
2SZ31NAK22B	Szoftver-technológia I.	K	gy							2	2	5												
4MA12NAK07B	Valószínűség-számítás	K	v							2	2	5												
2JO11NAK04B	Jog	K	v							2	2	5												
2SZ31NAK11B	Adatbázis rendszerek	K	v										2	2	5									
4OP13NAK01B	Operációkutatás	K	v										2	2	5									
4ST14NAK05B	Statisztika II.	K	v										2	2	5									
2SZ31NAK14B	Számítógép-hálózatok	K	v										2	2	5									
2SZ31NAK20B	Szoftver-technológia II.	K	gyj										2	2	5									
2SZ31NAK13B	Informatikai rendszerek fejlesztése	K	gyj													2	2	5						
2BE52NAK01B	Vállalati pénzügyek	K	v													2	2	5						
2SA53NAK01B	Számvitel alapjai	K	gyj													2	2	5						
2BE34NDK01B	Bevezetés az E-businessbe	K	v													2	2	5						
2VE81NGK03B	Szervezeti magatartás	K	v													2	0	3						
2VE81NGK07B	Menedzsment kontroll	K	v																2	1	5			
2IR32NDK11B	Minőség, audit	K	v																2	2	5			
2IR32NAK12B	Integrált rendszerek	K	v																2	2	5			
2EB34NAK01B	Médiagazdaságtan	K	v																2	2	5			
2IR32NAK11B	Infrastruktúra menedzsment	K	v																2	2	5			
2IR32NAK14B	Szakszeminárium	K	gyj																0	4	5			
Szakmai kötelező tárgyak																								20
2IR32NDK04B	Üzleti intelligencia	K	gyj																			1	2	5
2IR32NDK06B	Folyamat-menedzsment	K	gyj																			1	2	5
2IR32NDK15B	Szakszeminárium, diplomamunka	K	gyj																			0	4	10