

Hallgatói lemorzsolódás előrejelzésére szolgáló gépi tanulás modell tervezése

DOI: [10.29180/978-615-6342-50-8_6](https://doi.org/10.29180/978-615-6342-50-8_6)

SZERZŐK:

Fauszt Tibor¹, Erdélyi Katalin², Dobák Dóra³, Kovács Endre⁴

ABSZTRAKT

A hallgatói lemorzsolódás a felsőoktatási intézmények egyik fő problémája. Globális jelenség, hogy a beiratkozott hallgatók jelentős része nem jut el a diploma megszerzéséig. Az elsőévesek kiemelten veszélyeztetett csoportot alkotnak. Az első évben a követelményeket nem teljesítő hallgatókat nehéz a rendszerben tartani, így ők nagy valószínűséggel kiesnek, elhagyják a felsőoktatási intézményt. A lemorzsolódásnak számos oka és jelentős gazdasági, társadalmi hatása van, a hallgatók pedig személyes kudarcként élik meg. Az egyetemek kiemelt feladata, hogy erőfeszítéseket tegyenek a lemorzsolódásnak kitett hallgatók beazonosítására, és megfelelő, célzott segítségnyújtással azok rendszerben tartására. A gépi tanuláson alapuló, célzottan a hallgatói lemorzsolódásra megtervezett modellekkel a lemorzsolódás veszélyének kitett hallgatók megfelelő biztonsággal beazonosíthatók. Jelen tanulmányban a Budapesti Gazdasági Egyetem Pénzügyi és Számviteli Karán folyó Gazdaságinformatikus BSc képzés hallgatóinak az adminisztrációs rendszerből (Neptun) kinyert adatokon alapuló gépi tanulási modell építésével és előrejelző képességének vizsgálatával foglalkozunk.

Kulcsszavak: hallgatói lemorzsolódás, tanuláselemzés, gépi tanulás

ABSTRACT

Student dropout is one of the main problems of higher education institutions. It is a global phenomenon that a significant number of enrolled students do not get the diploma. First-year students are a particularly vulnerable group. It is difficult to keep students who do not meet the requirements in the first year in the system, so they are most likely to drop out and leave the higher education institution. Dropout has many reasons and significant economic and social effects, and students experience it as a personal failure. It is the priority task of universities to make efforts to identify students who are at risk of dropping out and to keep them in the system with appropriate, targeted assistance. With the models based on machine learning and specifically designed for student dropout, students at risk of dropping out can be identified with sufficient certainty. In this study, we deal with the construction of a machine

¹ Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, Gazdaságinformatika Tanszék, ORCID: 0000-0003-3590-1100, e-mail: fauszt.tibor@uni-bge.hu

² Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, ORCID: 0000-0001-5010-6688, e-mail: erdelyi.katalin@uni-bge.hu

³ Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, Gazdaságinformatika Tanszék, ORCID: 0000-0002-0857-9999, e-mail: dobak.dora@uni-bge.hu

⁴ Budapesti Gazdasági Egyetem, Pénzügyi és Számviteli Kar, Gazdaságinformatika Tanszék, ORCID: 0000-0001-6030-0716, e-mail: kovacs.endre@uni-bge.hu

learning model based on data extracted from the administration system (Neptun) and the examination of its predictive ability of the students of the Business Informatics BSc course at the Faculty of Finance and Accounting of the Budapest Business Scholl.

Keywords: student drop out, learning analytics, machine learning

Bevezető

A Gazdaságinformatikus BSc. alapképzés a Budapesti Gazdasági Egyetem Pénzügyi és Számviteli Karán a 2011-12- es tanévben indult 149 nappali tagozatos hallgatóval. A szak indulása óta töretlen népszerűségnek örvend a felvételizők körében. A létszámbővülés folyamatos, az utóbbi tanévekben minden évben hozzávetőlegesen 400 fő kezdi meg tanulmányait nappali és levelező képzésben, különböző finanszírozási formában. A képzésre felvett hallgatók nagyjából fele szerez diplomát, a másik fele lemorzsolódik, elhagyja az Egyetemet. Ez a lemorzsolódási arány nem kirívó, átlagosnak mondható a magyar felsőoktatásban (Vida & Németh, 2021; Miskolczi et. al., 2018). A lemorzsolódás mértéke a felsőoktatás egy minőségi indikátora, a felsőoktatási intézmények kiemelt feladata és érdeke is egyben, hogy lépéseket tegyenek a lemorzsolódási arány csökkentésére (Szemerszki, 2018). Az egyetemek által a lemorzsolódás csökkentésére tett intézkedések azonban nem mehetnek a minőség rovására. A hallgatói teljesítmény a sikeresség döntő tényezője, alapvető fontossággal bír. A képzési és kimeneti követelményeket mindenekelőtt szemmel tartva kell a tanulási folyamatba beavatkozni és lépéseket tenni a lemorzsolódás csökkentésére.

A felsőoktatási lemorzsolódást, mint fogalmat nehéz definiálni, annak bizonytalansága végigkíséri a kutatásokat (Pusztai & Szigeti, 2018). Általános értelemben, azokban az esetekben használjuk, amikor a hallgató végzettség nélkül hagyja el a felsőoktatást a saját, vagy az intézmény döntése alapján (Demcsákné & Huszárik, 2020). Azonban a képzés elhagyása nem feltétlenül jelenti azt, hogy a hallgató befejezi felsőoktatási tanulmányait, hisz előfordulhat, hogy más intézményt, más képzést választ, esetleg külföldön folytatja felsőoktatási tanulmányait. Jelen tanulmányban lemorzsolódáson a Gazdaságinformatika szakot végzettség nélkül elhagyó eseteket értjük.

A hallgatói lemorzsolódásnak számos kiváltó oka lehet, pl. családi háttér, rossz baráti társaság, munkavállalás, párkapcsolat, motiváció hiánya, képességbeli hiányosságok, korábbi tanulmányi kudarcok, stressz, elköteleződés hiánya, önértékelési problémák, rossz tanulmányi teljesítmény, pénzügyi nehézség, váratlan külső problémák, hogy hogyan érzi magát a diák a tanórákon, elkötelezettnek látja-e tanárait, érdekesnek-e az anyagot, és számos egyéb tényező. (Nonis & Fenner, 2012; Benett, 2003; Miskolczi et. al., 2018; Csók et. al., 2018; Szemerszki et. al., 2018; Demcsákné et. al., 2020; Vida & Németh, 2021; Dalipi et. al., 2018) Az itt felsorolt tényezők indikátorai lehetnek a lemorzsolódásnak, jelezhetik, hogy egy hallgató veszélyben van, azonban a sok esetben kényes adatok gyűjtése problematikus lehet, hiszen nem, vagy nehezen lehet adatot gyűjteni a lehetséges kiváltó tényezőkről (pl. szülők iskolai végzettsége, pénzügyi háttér) (GDPR). Továbbá bizonyos, a lemorzsolódást kifejező indikátort nehéz mérni (pl. motiváció, elkötelezettség). Felsorolt tényezők önmagukban nem feltétlenül vezetnek lemorzsolódáshoz, a jelenség nagyon összetett.

A lemorzsolódás további nagyon fontos indikátora a hallgatói teljesítmény (Benett, 2003). Számos hazai és nemzetközi tanulmány eredménye azt mutatja, hogy azon hallgatók, akik már az első félévekben sem nyújtanak kielégítő teljesítményt, nagy valószínűséggel diplomát sem szereznek, így

különösen veszélyeztetettek a tanulmányaikat kezdő elsőévesek. (Delen, 2010; Hoffait & Schyns, 2017; Ortiz & Dehon, 2013). A lemorzsolódás veszélyének kitett hallgatókat ezért a lehető legkorábban azonosítani kell. Amikor egy hallgató már meghozta döntését, miszerint idő előtt befejezi tanulmányait, már nehezen lehet a rendszerben tartani. Az azonosítás történhet közvetlenül a felvétel utáni időpontban, még a tanulmányok megkezdése előtt. A hallgatók beiratkozáskor rögzített adatait statisztikai eljárásokkal elemezve a lemorzsolódásnak veszélyének kitett hallgatók jó eséllyel azonosíthatók. Ahogy azonban már említettük, ezek az adatok sok esetben kényesek, nehéz az adatgyűjtés, vagy a mérés. Másik lehetőség a mielőbbi azonosításra, hogy a tanulmányok megkezdésével szinte egyidőben elindítjuk a hallgatói aktivitás, teljesítmény figyelemmel kísérését. Ennek egyik leghatékonyabb módja, ha egy adott LMS rendszerben (pl. Moodle) történik a tanulás és a rendszerbe épített eszközök (ML modellek) használatával azonosítjuk a lemaradó hallgatókat. A kurzusokat nem teljesítő hallgatók így már az első néhány hét után jó eséllyel azonosíthatók (Bognár et. al., 2021).

Gépi tanulási modellek a lemorzsolódás veszélyének kitett hallgatók beazonosítására

A felsőoktatás tömegessé válásával minden egyes hallgató teljesítményének nyomon-követése, a bukás, lemorzsolódás veszélyének kitett hallgatók beazonosítása informatikai rendszerek nélkül szinte lehetetlen feladat. A gépi tanulási módszerek, amelyek a mesterséges intelligencia részének tekinthetők, azonban segítséget nyújthatnak erre a problémára. (Kučak et. al., 2018; Tagharobi & Simbeck, 2022; Mi et. al., 2022; Ria et. al., 2022; Kovács et. al., 2022). Prediktív modellekkel hatékonyan azonosíthatók a lemorzsolódás veszélyének kitett hallgatók, azonban ez csakis jól megtervezett modellekkel lehetséges. Egy jó modell építéséhez nem elegendő egyszerűen csak a hallgatók adatait a modellbe táplálni, ez ennél sokkal összetettebb feladat, számos tényezőt figyelembe kell venni (Bognár & Fauszt, 2022). A modellek építésének alapfeltétele, hogy tömegesen álljanak az adatok rendelkezésre. Ezek után többek között: meg kell tervezni a célfüggvénnyel leginkább korreláló indikátorokat, meg kell határozni a leghatékonyabb modellszerkezetet, biztosítani kell továbbá, hogy a modellek ne legyenek érzékenyek az esetlegesen bekövetkező változásokra (pl. tantervi változások). Ezen felül számos tényezőt lehet még figyelembe venni, ami a tervezés során derülhet ki. A modellek építése előtt célszerű tanuláselemzést végezni, amelynek eredményeként a lemorzsolódás indikátorait feltárhatjuk. A tanuláselemzés (Learning Analytics, LA) a tanulókkal és környezetükkel kapcsolatos adatok gyűjtését és elemzését jelenti a tanulási folyamatok megértése és a tanulói eredmények javítása céljából. A tanuláselemzés (LA) egy multidiszciplináris terület, amely magába foglalja a gépi tanulást, a mesterséges intelligenciát, az információkeresést, statisztikát és adatvizualizációt (Chatti et. al., 2012).

Hallgatói teljesítmény és a lemorzsolódás összefüggése

A lemorzsolódás egy következmény, amelynek okait sok esetben nem tudjuk feltárni, viszont az előjeleit a hallgatói teljesítményben észlelhetjük. Az aktív tanulás, az egyenletes és jó teljesítmény nagyon fontos indikátora a kurzus sikeres teljesítésének (Freeman et. al., 2014). A kurzusok sikeres teljesítése pedig alapkövetelmény a diploma megszerzéséhez. A képzés, a felsőoktatás elhagyását fontolgató hallgató rosszul teljesíthet, ami megjelenhet tanulmányi eredményében. Ugyanakkor az a rosszul teljesítő hallgató is elveszítheti kedvét a továbbtanulástól, akinek a képzés elhagyására különösebb oka nincs, megfelelőek a körülményei a tanulásra, mégsem teljesít megfelelően.

Az eredményes tanulás és tanítás alapja a tantárgy teljesítéséhez szükséges követelmények pontos meghatározása (Farkas, 2019). A követelmények alapján tudjuk mérni a teljesítményt, amelyet általánosságban egy ötfokozatú skálán, érdemjegyekkel mérünk a magyar felsőoktatásban. A hallgatói teljesítményt kifejező adatok (érdemjegyek) gyűjtése, a teljesítményt kifejező indikátorok definiálása ebben az esetben egyszerű, hiszen a tanulmányi eredmények a hallgatói adminisztrációs rendszerben (Academic Management System) (Neptun) rendelkezésre állnak. A teljesítményt kifejező indikátorok használatával pedig jól megtervezett gépi tanulási modellek építhetők. Ezekkel a modellekkel a hallgatói lemorzsolódás előre jelezhető.

Modellépítés, adattisztítás, jellemzők meghatározása

A modellépítéshez a hallgatói adminisztrációs rendszerből (Neptun, Student Administration System) kinyert adatokat használtuk fel. Elsődlegesen azokkal a tömegével, minden hallgató esetében rendelkezésre álló adatokkal dolgoztunk, amelyek Demcsákné és Huszárik „Lemorzsolódási vizsgálatok a felsőoktatásban” c. tanulmánya alapján a lemorzsolódás hangsúlyos indikátorai. Ezek a tanulmány szerint a következők: hallgató demográfiai jellemzői, (neme, életkora), képzés jellemzői (passzív félévek száma, képzés munkarendje, finanszírozási forma), regionális jellemző (lakóhely régiója) (Demcsákné & Huszárik, 2020). Az adatokat a Neptun rendszerből 2011-12-1 félévtől 2020-2021-1 félévig kértük le, félévenkénti bontásban, Excel táblázatokban.

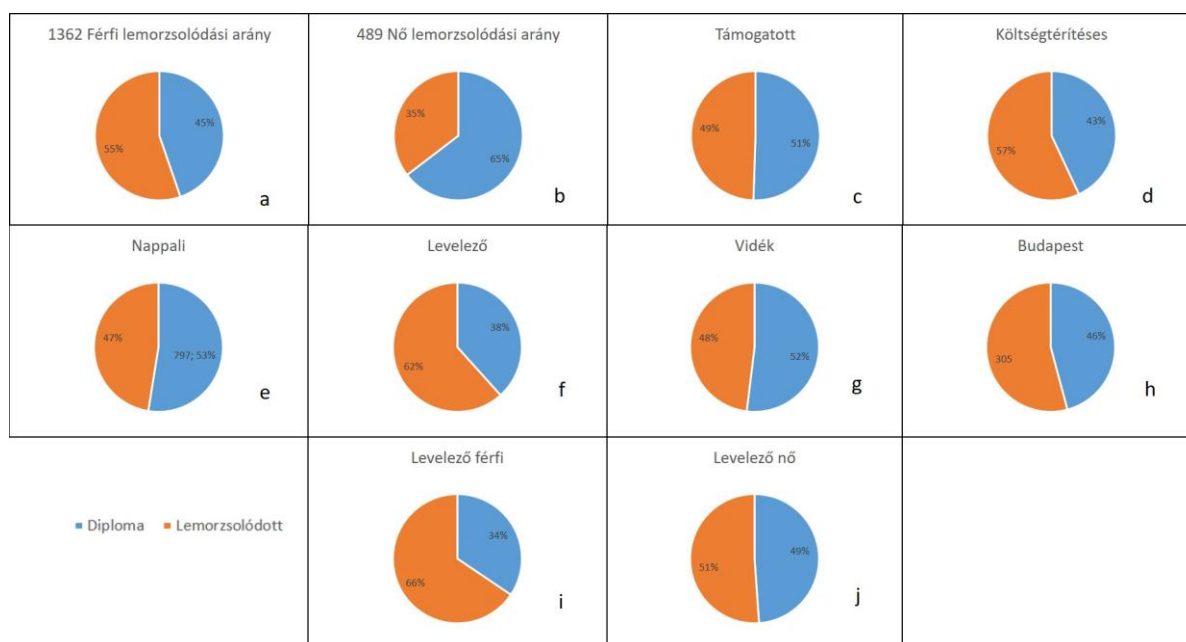
Az egyes táblázatok a következőket tartalmazták: születési idő, nem, lakóhely (város), tagozat, felvételi pontszám, felvétel dátuma, pénzügyi státusz (kölségtérítéses, önköltséges), aktív félévek száma, passzív félévek száma, státusz (van, nincs jogviszony) diploma eredménye. Ezt követően a hallgató által az adott félévben felvett tárgyak listája következett, amely a következő adatokat tartalmazta minden tárgy esetében: tárgykód, kredit, tárgyfelvelek száma, tárgyfelvelet féléve, tárgykövetelmény (aláírás, gyakorlati jegy, kollokvium), teljesítette-e, bejegyzés (aláírás, érdemjegy).

Az adatokat összesítettük, majd adattisztítást végeztünk. Első lépésben kiszűrtük azokat a hallgatókat, akik státusza inaktív, azaz megszűnt a hallgatói jogviszonyuk, elhagyták az egyetemet. Ezek után töröltük azokat a hallgatókat, ahol a felvétel dátuma többszörös értékkel rendelkezett. Ez a speciális eset akkor fordulhat elő, amikor a hallgató egy adott tanévben felvételt nyer, egy ideig aktív státuszú, teljesít néhány tantárgyat, majd leszámol, és újból felvételizik az egyetemre. Az újbóli sikeres felvételi után elfogadtatja az előzőleg teljesített tárgyait és folytatja a tanulmányokat. Ezek az esetek anomáliákhoz vezettek a modellben, ezért eltávolítottuk a listából őket. Összességében 1851 hallgató adatait kaptuk meg végül, mindegyikükről eldönthető volt, hogy melyik félévben nyert felvételt az egyetemre, továbbá, hogy szereztek-e diplomát, vagy sem. A vizsgálatban résztvevők lemorzsolódási aránya 49,97%-os, az 1851 hallgatóból 925 fő morzsolódott le.

Az adatok leíró statisztikai elemzése

A különböző szempont szerinti csoportok lemorzsolódási arányait a 1. sz. ábra mutatja. Az ábrák értelmezését az alábbiakban adjuk meg.

1. ábra Lemorzsolódási arányok különböző szempontok szerint



Forrás: saját kutatás, 2022, n=1851

Nemek és lemorzsolódási arány

A nemek között lényeges különbség látszik a lemorzsolódási arányok összevetése során. A férfiak és nők aránya az 1851 hallgató közül 1362 férfi és 489 nő. A férfiak körében a lemorzsolódás 55%-os, míg a nők esetében ez az arányt valamelyest kevesebbnek találtuk, 35%-os volt. 1/a és 1/b ábra.

Finanszírozási forma

A finanszírozási forma alapján a hallgatókat alapvetően két csoportra bontottuk, a valamilyen módon támogatott és a költségtérítéses képzésben résztvevőkre. Az 1/c és 1/d ábrán látható a két csoport lemorzsolódási aránya. A támogatott képzésben résztvevő hallgatók 49%-a morzsolódik le, míg a költségtérítéses képzéseknél ez az arány 57%. A saját erőből finanszírozott képzéseknél a hallgatói lemorzsolódás nagyobb, ugyanakkor nem számottevő a különbség.

Képzési forma

A 1/e és 1/f ábra a képzési forma (nappali, levelező) alapján csoportosított hallgatók lemorzsolódási arányát mutatja. A nappalis hallgatóknál a lemorzsolódási arány 47%, míg a levelezősöknél ez 62%. Látható tehát, hogy a levelezős hallgatók nagyobb arányban morzsolódnak le, mint a nappalisok. A levelezősök többsége munka mellett végzi a tanulmányait, ami extra terhelést jelent, hiszen sokkal kevesebb idő jut a hatékony tanulásra. Támogató családi háttér nélkül nehéz a levelező képzést elvégezni.

Regionális jellemző

Annak, hogy milyen szerepe van a lemorzsolódásnak a regionalitás szempontjából, a hallgatókat két csoportra bontottuk. Budapesti, illetve vidéki hallgatókra. A képzés a BGE PSZK-n folyik, amelynek campusa Budapesten található. A két kategória között nem találtunk lényeges különbséget, kis eltérés

mutatkozik a vidéki hallgatók csoportjának javára. A vidéken élő hallgatók esetében 48%-os, míg a budapesti hallgatók körében 54%-os lemorzsolódási arányt mutatnak az adatok (1/g és 1/h ábra).

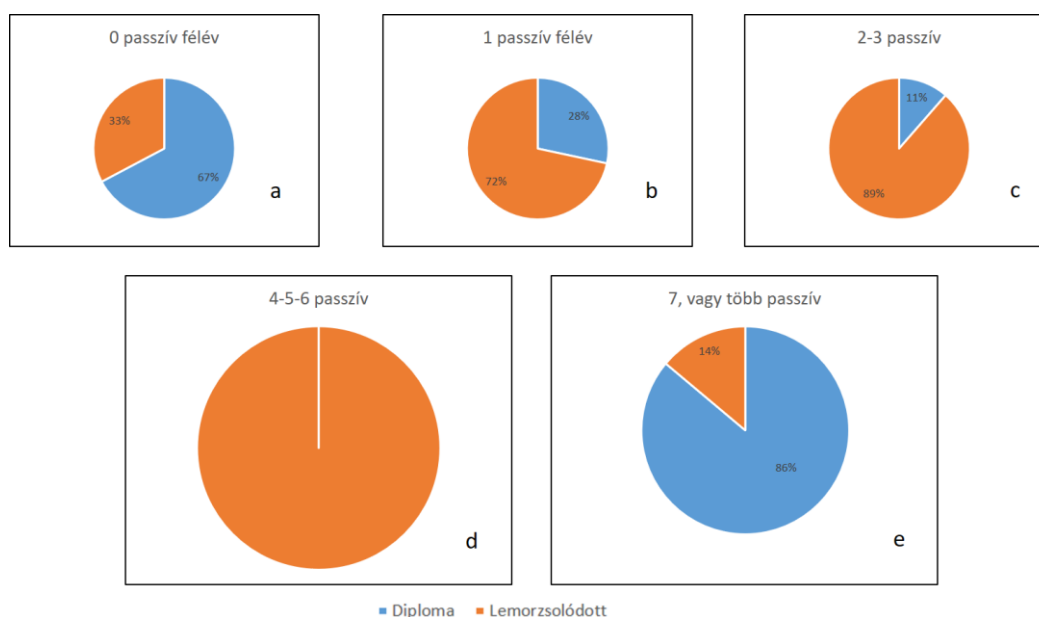
Férfi és nő lemorzsolódási arányok

Ebben a csoportosításban érdekes kép mutatkozik. A levelező férfiak lemorzsolódási aránya jelentősen magasabb, mint a levelező nőké. Az 1/i és 1/j ábrán látható, hogy a levelező férfiak mindössze 34%-a, míg a levelező nők 49%-a szerez diplomát (1/i és 1/j ábra). A levelező férfiak tehát a lemorzsolódás szempontjából kiemelt figyelmet igényelnek.

Passzív félévek száma és a lemorzsolódás összefüggése

Ebben a vizsgálatban azt néztük meg, hogy a passzív félévek száma és a lemorzsolódás milyen összefüggésben van egymással. A passzív félév azt jelenti, hogy a hallgató szünetelteti a tanulmányait, passzív félévre kerül, nem jár iskolába, nem vesz fel tantárgyat, nem vizsgázik. A hallgatókat 5 csoportba osztottuk. Az első csoportba azok kerültek, akik nem mentek passzív félévre egyetlen alkalommal sem. A következő csoportokba az 1, a 2-3, a 4-5-6, és a 7- vagy több passzív félévvel rendelkezők kerültek. Az adatokból az látszik, hogy a legkisebb a lemorzsolódási arány a 0, illetőleg a 7, vagy több passzív félévvel rendelkezők körében (2/a és 2/e ábra). A 0 passzív félévvel rendelkező hallgatók esetében a lemorzsolódási arány mindössze 33%-os, míg a 7- vagy több passzív félévvel rendelkezők esetében ez az arány még jobb, alig 14%-a morzsolódik le a hallgatóknak (2/a és 2/e ábra). Az 1 passzív félévvel rendelkezők lemorzsolódási aránya a 0 passzív félévvel rendelkezők lemorzsolódási arányával összehasonlítva 33%-ról 72%-ra, a 2-3 passzív félévvel rendelkezőkkel összehasonlítva 89% -ra nő (2/a és 2/b ábra). A 4-5-6 passzív félévvel rendelkezőknél a lemorzsolódás 100%-os, az ebbe a csoportba tartozó hallgatók egyike sem szerzett diplomát (2/d ábra). A passzív félévek száma tehát egy nagyon fontos tényező a lemorzsolódás szempontjából. Ezek az eredmények arra engednek következtetni, hogy a kitarató hallgatók, akik nem adják fel, akár 7, vagy több passzív félév után is diplomát szerezhetnek.

2. ábra Lemorzsolódási arányok a passzív félévek számának függvényében

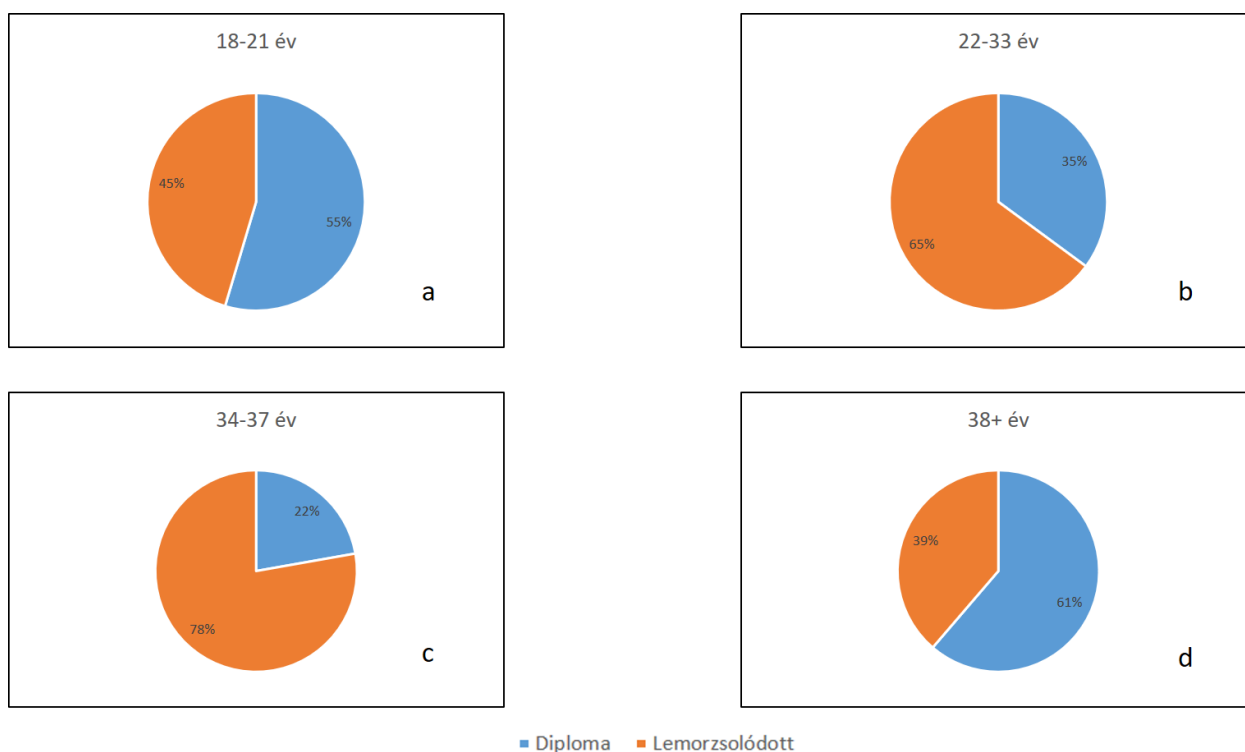


Forrás: saját kutatás, évszám, n=1851

Életkor és a lemorzsolódás

Az életkor alapján a hallgatókat négy csoportba osztottuk. Külön csoportba kerültek a 18-21, 22-23, 34-37 és a 38 és az a felletti életkorral rendelkező hallgatók. A legjobban a 18-21 és 38+ csoportban lévők teljesítenek, 55- és 61%-os diplomaszerzési aránnyal rendelkeznek (3/a és 3/d ábra). A 22-33 életkorral rendelkezők csoportjában lévő hallgatók mindössze 35%-a szerez diplomát. A legrosszabb teljesítményt a 34-37 éves hallgatók nyújtják, az ebbe a csoportba tartozók a legveszélyeztetettebbek, alig 22%-uk szerez diplomát (3/b és 3/c ábra). Információk hiányában a jelenség pontos hátterét nem elemeztük, azonban valószínűsíthető, hogy a családalapítás, a speciális élethelyzet az, ami a nagyarányú lemorzsolódást eredményezi.

3. ábra Lemorzsolódási arányok az életkor függvényében



Forrás: saját kutatás, 2022, n=1851

A gépi tanulási modell indikátorainak definiálása

Jelen kutatásunkban célunk az, hogy a beiratkozáskor rendelkezésre álló adatok és a hallgató beiratkozás utáni első két féléves teljesítménye alapján olyan prediktív modellt építsünk, amely nagy valószínűséggel megjósolja, hogy a hallgató ki van-e téve a lemorzsolódás veszélyének.

A modellben 16 indikátort definiáltunk, amelyek típusuk szerint alapvetően két csoportba sorolhatók. Az egyik típusba azok az indikátorok tartoznak, amelyek a beiratkozáskor rendelkezésre állnak. Ezek az indikátorok alapvetően a leíró statisztikai elemzésben felsorolt adatok alapján definiáltak. Ezek az indikátorok a következők:

1. *Tagozat* (nappali, levelező),
2. *Nem* (férfi, nő),
3. *Lakóhely* (Budapest, Vidék),
4. *Felvételi pontszám*,
5. *Pénzügyi státusz* (Támogatott, Költségtérítéses),
6. *Életkor*,

A következő csoportba azok az indikátorok tartoznak, amely indikátorok értéke a két félév során felvett tárgyak darabszámából, illetőleg azok teljesítéséből számíthatódnak. Mielőtt az egyes indikátorokat felsoroljuk, néhány fogalmat tisztázunk. Az alaptanterv követelményrendszere szerint a tárgyakat alapvetően három módon lehet teljesíteni. Követelmények: (a) aláírás, ahol a tárgy teljesítéséhez aláírás kell, értéke lehet aláírás, illetőleg aláírás megtagadva lehet, (b) gyakorlati jegy, ahol a tárgy teljesítéséhez gyakorlati jegyet kell szerezni a szorgalmi időszakban, értéke lehet 1, 2,...,5., (c) kollokvium, ahol a tárgy teljesítéséhez vizsgáznia kell a hallgatónak, ennek az értéke is szintén 1,2,...,5. Gyakorlati jegy és kollokvium követelmény esetén teljesített a tárgy akkor, ha legalább elégségest kapott a hallgató az adott tárgyból. Minden egyes tárgyhoz kreditérték van rendelve, ami a tárgy teljesítése esetén növeli az adott félévben teljesített összesített kreditértéket. Ennek alapján a következő, a hallgatói teljesítményt kifejező időfüggő indikátorokat definiáltuk a modellhez az első két félévre vonatkoztatva:

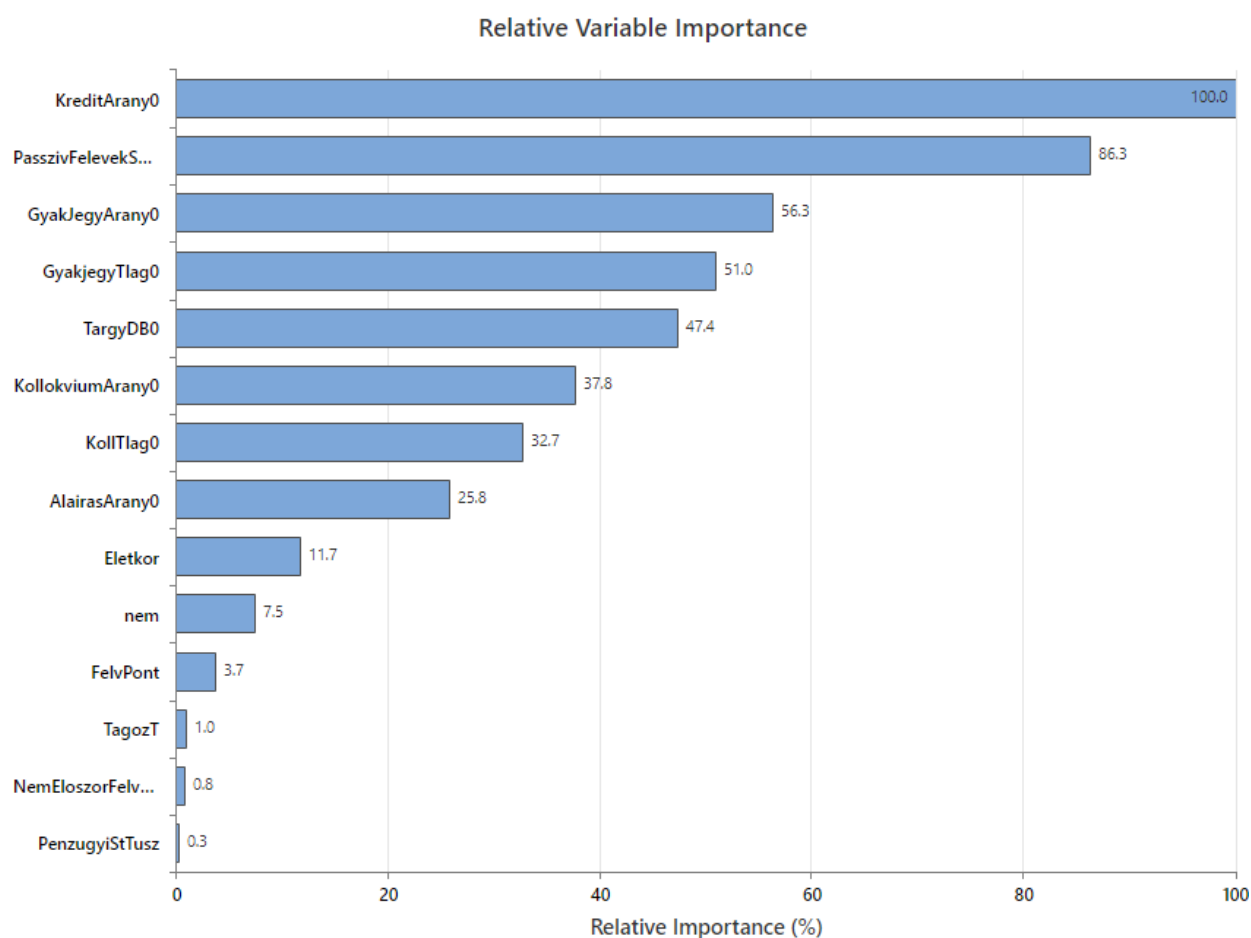
7. *Felvett tárgyak száma*,
8. *Aláírás aránya* (Megszerzett aláírások száma aláírás követelményű tárgyak esetén / Aláírás követelményű tárgyak száma),
9. *Kollokvium aránya* (Legalább elégséges szinten teljesített kollokvium követelményű tárgyak száma / Kollokvium követelményű tárgyak száma),
10. *Kollokvium átlag* (Kollokvium követelményű tárgyak átlaga),
11. *Gyakorlat aránya* (Legalább elégséges szinten teljesített gyakorlati jegy követelményű tárgyak száma / Gyakorlati jegy követelményű tárgyak száma),
12. *Gyakorlat átlag* (Gyakorlati jegy követelményű tárgyak átlaga),
13. *Kredit arány*, (Teljesített tárgyak kreditösszege / Felvett tárgyak kreditösszege)
14. *Nem először felvett tárgyak száma*,
15. *Nem először felvett teljesített tárgyak száma*,
16. *Passzív félévek száma*.

A modellben összesen 1851 hallgatóhoz tartozó fenti 16 indikátorérték szerepelt két félév adataira számítva.

4. Erdemények

Az egyes prediktorok relatív fontosságát a 4. sz. ábrán láthatjuk.

4. ábra *Prediktorok relatív fontossága*



Forrás: saját kutatás, 2022

A lemorzsolódás szempontjából a legfontosabb indikátornak a *Kreditarány* bizonyult. Ez az indikátor azt fejezte ki, hogy a felvett tárgyakat milyen arányban teljesítette sikeresen a hallgató. A következő, igen fontos indikátor a *Passzív félévek száma*, ami a leíró statisztika eredményeiből is várható volt. A *Gyakorlati jegy arány*, *Gyakorlati jegy átlag* a *Felvett tárgyak száma*, a *Kollokvium arány* és *Kollokvium átlag* indikátorok fontossága is számottevő, de messze nem annyira, mint a *Kreditarány*, vagy a *Passzív félévek száma* indikátoroké. Ez arra enged következtetni, hogy azok a hallgatók, akiknek a tanulmányi eredményük nem kiemelkedő, nem szereznek jó jegyeket, hasonló eséllyel szereznek diplomát, mint azok a hallgatók, akiknek a tanulmányi eredménye kimagasló, jó eredményeket érnek el a vizsgákon. További relatív fontos indikátoroknak tekinthetők az *Életkor*, és a *Nem*. A *Felvételi pontszám*, *Tagozat*, a *Nem először felvett tárgyak száma*, a *Nem először felvett tárgyak teljesítési aránya* relatíve alacsony fontossággal, a *Lakóhely* és a *Pénzügyi státusz* pedig elenyésző fontossággal bír a modellben. A *Felvett tárgyak száma* indikátor relatíve magas fontossága összefüggésben lehet a tanulmányi előmenetellel. Az első félévben kevés tárgyat teljesítő hallgatók a következő félévben több tárgyat vettek fel, hogy tanulmányi kötelezettségeiket teljesítsék. Ez azonban

sok hallgató esetében oda vezetett, hogy nem tudták teljesíteni a követelményeket, a sok teljesítendő tárgy olyan nyomás, amit kevesen bírtak.

A modellt a Matlab programban hoztuk létre, a modell gépi tanulási algoritmusául a döntési fát választottuk. A döntési fa egy felügyelt gépi tanulási algoritmus, amely leginkább klasszifikációs problémákra alkalmas. Esetünkben ez az algoritmus bizonyult a legjobbnak a lemorzsolódás előrejelzésére. A betanított modellt tesztadatokra vonatkozó előrejelzésekre használtuk. A tesztadatokon végzett előrejelzésekből generált tévesztési mátrix az 1. számú táblázatban látható.

5. ábra Tévesztési mátrix

Confusion Matrix								
Actual Class	Predicted Class (Training)				Predicted Class (Test)			
	Count	0	1	% Correct	Count	0	1	% Correct
0 (Event)	651	546	105	83.9	275	234	41	85.1
1	646	59	587	90.9	279	26	253	90.7
All	1297	605	692	87.4	554	260	294	87.9
Statistics				Training (%)	Test (%)			
True positive rate (sensitivity or power)				83.9	85.1			
False positive rate (type I error)				9.1	9.3			
False negative rate (type II error)				16.1	14.9			
True negative rate (specificity)				90.9	90.7			

Forrás: saját kutatás

A képzési adatokra épített döntési fa modell teszt adatokra alkalmazott klasszifikációs pontossága 87,9%. A modellt elsődlegesen azokra az esetek előrejelzésére építettük, ahol a hallgató várhatóan lemorzsolódik, nem szerez diplomát. Ez a modellben a 0 esemény (0 Event). A modell érzékenysége, (True positive rate, sensitivity) a tesztesetekre vonatkoztatva 85,1%, ami azt jelenti, hogy azokat a hallgatókat, akik nem szereztek diplomát a modell 85%-os valószínűséggel azonosítja. Azokat a hallgatókat, akiket abba a csoportba sorol, akik várhatóan diplomát szereznek (True negative rate, specificity) és a valóságban is diplomát szereztek, 90,7% pontossággal azonosította. Az elsőfajú hiba (False positive rate), azt jelenti, hogy a modell hibásan azt jósolja, hogy a hallgató nem szerez diplomát, miközben a valóságban szerzett. Ennek a mutatónak az értéke 9,3%. A másodfajú hiba (False negative rate), azt jelenti, hogy a modell hibásan azt jósolja, hogy a hallgató szerez diplomát, miközben nem szerzett. Ennek a mutatónak az értéke 14,9%. Ezek a mutatók elfogadható értékkel bírnak, a modell nagy valószínűséggel azonosítja a lemorzsolódás veszélyének kitett hallgatókat azonban az esetek 15%-ában nem tudja azonosítani azokat, akik a valóságban lemorzsolódtak. A mutatók tovább javíthatók, ha 2-nél több félév adatait vesszük figyelembe, azonban ahogy a bevezetőben említettük a korai azonosítás alapvetően fontos a lemorzsolódás kezelésében, megfékezésében.

Összefoglalás

A mesterséges intelligencia, a gépi tanulási eszközök az utóbbi években hihetetlen fejlődésen mentek keresztül, mindennapi életünk elengedhetetlen eszközévé kezdenek válni. Ez a technológia tért hódít

az autóiparban, a képfeldolgozásban, a szociális médiában, a kereskedelemben, az orvostudományban, az oktatásban, és számos egyéb területen. Ezt a technológiát elsősorban olyan környezetben használhatjuk hatékonyan, ahol a körülmények állandónak tekinthetők. Változó környezet esetében, mint például az oktatás, gondos tervezést és elemzést igényel a technológia alkalmazása. Hiszen egy oktatási folyamatban az évek során sok minden változhat. Változhat a tanterv, az oktató, a hallgatók előképzettsége, tantárgyi tartalom, a félévekben felvett tárgyak száma, és sok egyéb más tényező. Erre a változó környezetre prediktív gépi tanulási modell csak úgy építhető, ha olyan prediktorokat építünk, amelyek ebben a változó környezetben is jól teljesítenek, szoros összefüggésben vannak a célfüggvénnyel. A vizsgálatunkban épített modellben biztosítani tudtuk, hogy a változó környezet ne befolyásolja a modell prediktív képességét, ami végső soron a modell jósági mutatóiban nyilvánult meg. Az általunk épített gépi tanulási modell elfogadható pontossággal azonosítja a lemorzsolódás veszélyének kitett hallgatókat. A hallgatói adminisztrációs rendszer adataira épülően megfelelő informatikai fejlesztésekkel kiegészítve hatékony eszköz lehet képzésfejlesztés területén, elsődlegesen az egyetem hallgatói tanulás támogató tevékenységet illetően.

Irodalomjegyzék

1. Bennett, R. (2003). Determinants of undergraduate student drop out rates in a university business studies department. *Journal of further and Higher Education*, 27(2), 123-141.
2. Bognár, L., & Fauszt, T. (2022). Factors and conditions that affect the goodness of machine learning models for predicting the success of learning. *Computers and Education: Artificial Intelligence*, 3, 100100.
3. Bognár, L., Fauszt, T., & Nagy, B. (2021). Machine Learning Model Building Techniques for Small and Medium-sized University Courses. *International Journal of Artificial Intelligence*, 19(2), 20-43.
4. Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International journal of Technology Enhanced learning*, 4(5-6), 318-331.
5. Csók, C., Dusa, Á. R., Hrabéczy, A., Novák, I., Karászi, Z., Ludescher, G., & Németh, D.. (2018). Ahallgatói lemorzsolódás és háttértényezői egy kvalitatív kutatás tükrében. *Lemorzsolódás és perzisztencia a felsőoktatásban*, 38-62.
6. Dalipi, F., Imran, A. S., & Kastrati, Z. (2018, April). MOOC dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE global engineering education conference (EDUCON)* (pp. 1007-1014). IEEE.
7. Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), pp. 498-506.
8. Demcsákné, O. Zs., Huszárík, P., (2020). Lemorzsolódási vizsgálatok a felsőoktatásban. Összefoglaló tanulmány. EFOP-3.4.5-VEKOP-17-2017-00001. Rendszerszintű fejlesztések és hozzáférés bővítését szolgáló ágazati programok a felsőoktatásban. *Oktatási Hivatal kiadványa*, 2020.

9. Farkas, É. (2019). Tanulási eredmények értékelése a felsőoktatásban. *University of Szeged, Szeged*.
10. Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences, 111(23)*, 8410-8415.
11. Pusztai, G. & Szigeti, F. (Eds.), (2018) Lemorzsolódás és perzisztencia a felsőoktatásban [Dropout and persistence in higher education] (pp. 38–62). *Debrecen, Hungary: Debreceni Egyetemi Kiadó*.
12. Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems, 101*, 1-11.
13. Kovács, G., Erdélyi, K., Pádár, Z., Zenke, P., & Nogel, M. (2022). Contamination of forensic DNA evidence in the light of Hungarian court decisions: A review of 25 years. *Forensic Science International: Genetics Supplement Series*.
14. Kučak, D., Juričić, V., & Đambić, G. (2018). Machine Learning In Education. A Survey of Current Research Trends. *Annals of DAAAM & Proceedings, 29*.
15. Mi, H., Gao, Z., Zhang, Q., & Zheng, Y. (2022). Research on Constructing Online Learning Performance Prediction Model Combining Feature Selection and Neural Network. *International Journal of Emerging Technologies in Learning, 17(7)*.
16. Miskolczi, P., Bársony, F., & Király, G. (2018). Hallgatói lemorzsolódás a felsőoktatásban: elméleti, magyarázati utak és kutatási eredmények összefoglalása. *Iskolakultúra, 28(3-4)*, 87-105.
17. Nonis, S. A., & Fenner, G. H. (2012). An Exploratory Study of Student Motivations for Taking Online Courses and Learning Outcomes. *Journal of Instructional Pedagogies, 7*.
18. Ria, A., Harry, B. D., Zaniyal, A. H., (2022). Monitoring Learners' Performance By Modelling Learning Process Using Machine Learning. *Journal of Engineering Science and Technology Special Issue on SIMASE, August (2021) 30 – 39*.
19. Tagharobi, H., & Simbeck, K. (2022). Introducing a Framework for Code based Fairness Audits of Learning Analytics Systems on the Example of Moodle Learning Analytics. *In CSEDU (2) (pp. 45-55)*.

20. Szemerszki, M. (2018). Lemorzsolódási adatok es módszertani megfontolások. In G. Pusztai & F. Szigeti (Eds.), *Lemorzsolódás és perzisztencia a felsőoktatásban [Dropout and persistence in higher education]* (pp. 15–27). *Debrecen, Hungary: Debreceni Egyetemi Kiadó.*
21. Vida, C., Németh, E. (2021). Felsőoktatás a változások tükrében –verseny, minőség, teljesítmény. *Magyar Akkreditációs Szemle*. 2. szám. pp.: 21-25.