

Lemorzsolódás előrejelzése személyre szabott értelmezhető gépi tanulási módszerek segítségével

Nagy Marcell 

Budapesti Műszaki és Gazdaságtudományi Egyetem, Matematika Intézet, Sztochasztika Tanszék,
Budapest, Magyarország
eKRÉTA Informatikai Zrt., Budapest, Magyarország
E-mail: nagy.marcell@ttk.bme.hu

Beérkezett: 2022. augusztus 22.; elfogadva: 2022. november 14.

Összefoglalás

A hallgatói lemorzsolódás az egyik legégetőbb probléma a felsőoktatásban. Ebben a munkában a lemorzsolódás előrejelzésén keresztül bemutatjuk, hogyan tudják segíteni a felsőoktatás résztvevőit a magyarázható mesterséges intelligencia (XAI) eszközök, mint például a permutációs fontosság, a parciális függőségi ábra és a SHAP. Végül pedig kitérünk a kutatás gyakorlati hasznosulásának lehetőségeire, például, hogy az egyéni előrejelzések magyarázata hogyan teszi lehetővé a személyre szabott beavatkozást. Az elemzések során azt találtuk, hogy a középiskolai tanulmányi átlag bír a legnagyobb prediktív erővel a végzés tényére vonatkozóan. Továbbá annak ellenére, hogy egy műszaki egyetem adatait elemeztük, azt találtuk, hogy a humán tárgyaknak is nagy inkrementális prediktív erejük van a végzés tényére vonatkozóan a reál tárgyakhoz képest.

Kulcsszavak: lemorzsolódás előrejelzése, felsőoktatás, oktatási adattudomány, értelmezhető gépi tanulás, megmagyarázható mesterséges intelligencia

Predicting dropout using personalized interpretable machine learning tools

Marcell Nagy

Budapest University of Technology and Economics, Institute of Mathematics, Department of Stochastics,
Budapest, Hungary
eKRÉTA Infromatics, Budapest, Hungary

Summary

Delayed completion and student drop-out are some of the most critical problems in higher education, especially regarding STEM programs. A high drop-out rate induces both individual and economic loss, hence a detailed investigation of the main reasons for dropping out is warranted. Recently, there has been a lot of interest in the use of machine learning methods for the early detection of students at risk of dropping out. However, there has not been much debate on the use of interpretable machine learning (IML) and explainable artificial intelligence (XAI) technologies for dropout prediction. In this paper, we show how IML and XAI techniques can assist educational stakeholders in dropout prediction using data from the Budapest University of Technology and Economics. We demonstrate that complex black-box machine learning algorithms, for example CatBoost, are able to effectively detect at-risk student using only pre-enrollment achievement measures, but they lack interpretability. We demonstrate how the predictions can be explained both globally and locally using IML methods including permutation importance (PI), partial dependence plot (PDP), LIME, and SHAP values.

Using global interpretations, we have found that the factor that has the greatest impact on academic performance is the high school grade point average, which measures general knowledge by taking into account grades in history, mathematics, Hungarian language and literature, a foreign language and a science subject. However, we also found that both mathematics and the subject of choice are among the most important variables, which suggests that pro-

gram-specific knowledge is not negligible and complements general knowledge. We discovered that students are more likely to drop out if they do not start their university studies immediately after leaving secondary school. Using a partial dependence plot, we showed that humanities also have incremental predictive power, despite the fact that this analysis is based on data from a technical university.

Finally, we also discuss the potential practical applications of our work, such as how the explanation of individual predictions allows for personalized interventions, for example by offering appropriate remedial courses and tutoring sessions. Our approach is unique in that we not only estimate the probability of dropping out, but also interpret the model and provide explanations for each prediction. As a result, this framework can be used in several fields. By predicting which majors they could be most successful in based on high school performance indicators, it might, for instance, assist high school students in selecting the appropriate programs at universities and hence this way it could be used for career assistance. Through the explanations of local predictions, the framework provided can also assist students in identifying the skills they need to develop to succeed in their university studies.

Keywords: dropout prediction, higher education, interpretable machine learning, educational data science, explainable AI

Előszó

A Kooperatív Doktori Program keretében Nagy Marcell eKRÉTA-ban végzett legfőbb feladata, hogy a Budapesti Műszaki és Gazdaságtudományi Egyetem kutatási előzményeire alapozva többintézményes tanulmányi adatokon a gépi tanulás modern eszközeit alkalmazva kutatás-fejlesztést végezzen. A kutatás gyakorlati eredménye olyan mesterségesintelligencia-alapú döntéstámogató eszköz/alkalmazás, amely hasznos mind a középiskolások, mind az egyetemisták, mind az oktatási döntéshozók számára.

A KDP kutatás gyakorlati alkalmazásának irányába komoly előrehaladást értünk el, amiben Nagy Marcell komoly szerepet játszott. Az elmúlt másfél évben egy olyan termék mesterséges intelligenciájának fejlesztésére fókuszáltunk, ami képes egy diák középiskolai teljesítményei alapján előre jelezni az egyetemi sikerességét vagy a lemorzsolódás veszélyét, megállapítva az ehhez hozzájáruló tényezőket és figyelembe véve a képzési területet is.

A jelen közleményben bemutatott legfrissebb kutatási eredmények az értelmezhető gépi tanulási eszközök alkalmazásáról a lemorzsolódás előrejelzésében közvetlenül kerülnek majd felhasználásra és alkalmazásra a következő időszakban. Az értelmezhető gépi tanulási eszközök segítik majd a felhasználókat megérteni, interpretálni a modell előrejelzését, illetve a modul döntéstámogató része közvetlenül ezeken a megmagyarázható mesterségesintelligencia-eszközökön alapul.

*Dr. Szabó Balázs
vállalati szakértő, vezérigazgató
eKRÉTA Informatikai Zrt.*

Nagy Marcell óriási lendülettel vetette bele magát a Kooperatív Doktori Program keretében zajló kutatás-fejlesztési projekt megvalósításába az eKRÉTA Informatikai Zrt.-vel való együttműködésben. A 2021 februárja óta tartó ösztöndíjas időszak alatt Marcell példaértékű tudományos produktivitásról tett tanúbizonyságot. Ez alatt az időszak alatt Marcellnek nyolc tudományos közleménye jelent meg, amelyből hat Q1-es folyóiratban lett publikálva. Mindemellett a tudományos eredményeinek gyakorlati

alkalmazásba történő átültetésében is komoly előrelépéseket ért el.

Ebben a tanulmányban Nagy Marcell legújabb kutatási eredményei kerülnek bemutatásra, ahol különböző értelmezhető gépi tanulási módszerek alkalmazhatóságát vizsgálja a lemorzsolódás előrejelzésével kapcsolatban. Ez a munka amellett, hogy innovatív, olyan, az oktatási adattudományban legmodernebb technikákat alkalmazza, amelyek mély matematikai modelleken alapulnak. Például, a mű egyik fontos eleme az úgynevezett SHAP-értékek, amelyek a játékelméleti Shapley-érték fogalmán alapulnak.

Fontos kiemelni, hogy Nagy Marcell témavezetésébe konzulensként Molontay Roland kollégám is bekapcsolódott, akivel hét közös tudományos közleménye jelent meg oktatási adattudományi témában.

*Dr. Simon Károly
témavezető, tanszékvezető egyetemi tanár
BME Sztochasztika Tanszék*

Bevezetés

Az innovatív és hatékonyan működő felsőoktatás az alapja a sikeres tudás- és technológiaalapú gazdaságnak, ahol a szolgáltatások és a termelés tudásintenzív és szellemi tevékenységekre támaszkodik (Powell–Snellman 2004). A diákok lemorzsolódása vagy a tanulmányok elhúzódo befizetése azonban világszerte komoly problémát jelent a felsőoktatásban, különösen a STEM/MTMI (matematikai, természettudományi, mérnöki és informatikai) képzések esetén, súlyos személyes és társadalmi költségeket maga után vonva (Latif–Choudhary–Hamayun 2015).

Az utóbbi időben számos kutatás született, amelyben prediktív analitikai módszereket alkalmaztak a lemorzsolódásban veszélyeztetett hallgatók azonosítására vagy az egyetemi sikerességre ható tényezők meghatározására (Márquez-Vera et al. 2016; Rovira–Puertas–Iguar 2017; Helal et al. 2019; Alyahyan–Düştögör 2020; Varga–Sátán 2021). Az elmúlt években a felsőoktatási szereplők segítségével több mesterségesintelligencia-alapú döntéstámo-

gató rendszert is fejlesztettek világszerte, ezekről az alábbi cikkek adnak jó áttekintést: *Avella et al. 2016; Dutt–Ismail–Herawan 2017; Zawacki–Richter et al. 2019; Rastrollo–Guerrero–Gomez–Pulido–Durán–Domínguez et al. 2020.*

A lemorzsolódást előrejelző tanulmányok túlnyomó többségének az a célja, hogy gépi tanulási algoritmusokkal minél előbb, minél pontosabban meg tudják különböztetni (osztályozni) a lemorzsolódásban veszélyeztetetteket és a sikeresen diplomázókat. Azonban ezekben a munkákban a betanított modellek értelmezése és az előrejelzéseik megmagyarázása általában elmarad. Egy jól működő döntéstámogató rendszer mellett, hogy nagy hatékonysággal azonosítani tudja a veszélyeztetett diákokat, képes magyarázatot is adni az előrejelzésre, ezáltal segítve az érintetteket abban, hogy különböző intervenciókkal vagy változtatásokkal növelni tudják a diploma-szerzés esélyét. Például egy jól interpretálható előrejelzés segítségével személyre szabott útmutatást, felzárkóztató kurzusokat és korrepetálásokat ajánlhatunk a lemorzsolódásban veszélyeztetett hallgatók számára. Ezen túlmenően az előrejelzések magyarázata és az eredmények értelmezése segít a modell átláthatóbbá tételében és a döntéshozók bizalmának megteremtésében is, ami kulcsfontosságú lépés e megoldások valós oktatási környezetben való telepítéséhez és alkalmazásához. Ezek a célok könnyen elérhetőek a gépi tanulás egy nemrégiben kialakult területének, az értelmezhető gépi tanulásnak (IML) vagy más néven magyarázható mesterséges intelligenciának (XAI) az eszközeinek segítségével. Ennek az új területnek a célja olyan módszerek kutatása, amelyek segítségével átláthatóvá és értelmezhetővé tehetők a legújabb komplex, eredendően nem értelmezhető ún. „fekete doboz” (black box) gépi tanulási modellek is. Az IML eszközeiről alapos átfogó áttekintését nyújt Molnar könyve (*Molnar 2020*) és Adadi és Berrada munkája (*Adadi–Berrada 2018*).

Molnar (2020) taxonómiája szerint a gépi tanulási algoritmusokat értelmező módszerek a következők szerint csoportosíthatók: (1) intrinzik (belső/valódi/természetes) vagy post hoc (utólagos), (2) modellspecifikus vagy modell-agnosztikus, és (3) lokális vagy globális. Az intrinzik módszerek olyan alacsony komplexitású ún. „white-box” modelleket foglalnak magukban, amelyek IML-eszközök használata nélkül értelmezhetőek, mint például a lineáris regresszió és a döntési szabályok (pl. döntési fák). Ezzel ellentétben a post hoc módszerek már egy betanított modellt elemeznek utólag. A modellspecifikus értelmezés azt jelenti, hogy ismerni kell a modell mechanizmusát ahhoz, hogy értelmezni tudjuk az előrejelzését: a belsőleg értelmezhető modellek értelmezése mindig modellspecifikus. A modell-agnosztikus eszközök bármely modellre használhatók a betanítás után, azaz a modell-agnosztikus eszközök mindig post hoc eszközök. Végül azt mondjuk, hogy az értelmezés lokális, ha egy egyéni előrejelzést magyaráz meg, és globális, ha a teljes modell viselkedését írja le.

Számos cikk használ belsőleg értelmezhető módszereket az oktatási adattudományban, például Márquez-Vera és munkatársai (*Márquez-Vera et al. 2016*) az ún. Interpretable Classification Rule Mining (ICRM – értelmezhető osztályozási szabály bányászat) algoritmus egy módosított változatát alkalmazták, hogy a lehető legkorábban azonosítsák a lemorzsolódásban veszélyeztetett diákokat egy kurzuson. Hasonlóképpen Zhang, Zhou és Yi is hangsúlyozzák (*Zhang–Zhou–Yi 2019*) az értelmezhetőség fontosságát, és online tanulók teljesítményének előrejelzésére ők is az ICRM algoritmust alkalmazzák. Cano és Leonard az alulreprezentált és aluteljesítő diákokra összpontosítanak (*Cano–Leonard 2019*), és egy értelmezhető korai figyelmeztető rendszert vezetnek be, amely a többnézetes genetikai programozásra (multiview genetic programming) támaszkodik, és a szabálygeneráló (rule mining) algoritmushoz hasonló döntési szabályokat hoz létre. Az online tanulási rendszerekben a lemorzsolódás kockázatának előrejelzésére Coussement és munkatársai (*Coussement et al. 2020*) a logit leaf modellt (LLM) használják, amely egy döntési fát és logisztikus regressziót kombinál, és azt állítják, hogy az LLM remek egyensúlyt biztosít a jó előrejelző teljesítmény és az értelmezhetőség között. Ezen túlmenően a szerzők egy többszintű informatív vizualizációt is fejlesztettek az LLM-hez, amely nemcsak a lemorzsolódás valószínűségét mutatja, hanem segít megérteni a modell lokális, egyéni szintű előrejelzéseit is.

Kevés olyan tanulmány létezik, amely post hoc modell-agnosztikus értelmezhető gépi tanulási eszközöket (pl. LIME és SHAP) alkalmaz oktatási környezetben. Nagrecha és munkatársai online kurzusokon (MOOC) való lemorzsolódást jelezték előre (*Nagrecha–Dillon–Chawla 2017*), és a LIME algoritmus segítségével tették értelmezhetővé a „fekete doboz” gépi tanulási modelleket (Random Forest és Gradient Boosting Trees), továbbá összehasonlítást is végeztek a belsőleg értelmezhető modellekkel (logisztikus regresszió és döntési fa). Nemrégiben *Vultureanu–Albiși és Bădică (2021)* faalapú együttes (ensemble) osztályozókat használt a hallgatók kurzusokon nyújtott teljesítményének előrejelzésére nyilvánosan elérhető adathalmazok alapján, továbbá Nagrecha és munkatársaihoz hasonlóan ők is alkalmazták a LIME algoritmust a modell lokális előrejelzéseinek értelmezésére (*Nagrecha–Dillon–Chawla 2017*). Sargsyan és munkatársai (*Sargsyan et al. 2020*) klaszterelemzést végeztek a hasonló jellemzőkkel rendelkező diákok csoportjainak megtalálására. Először a klaszterezést a diákok előre jelzett tanulmányi átlaga alapján végezték el, majd a LIME által adott lokális előrejelzések magyarázatából származó súlyok alapján. A szerzők azzal érvelnek, hogy az általuk javasolt megközelítés olyan diákcsoportokat talál, amelyek hasonló tanulmányi eredménymutatókkal rendelkeznek.

A legtöbb kapcsolódó munka a LIME algoritmust használja a lokális előrejelzések magyarázatára, azonban néhány frissebb tanulmányban a Shapley Additive Expla-

nations (SHAP) játékelmélet-alapú módszert alkalmazták, amelyet Lundberg és Lee vezetett be (Lundberg–Lee 2017). Mingyu és munkatársai az egyetemi hallgatók súlyozott tanulmányi átlagát jelezték előre egy CatBoost regresszor segítségével, továbbá a SHAP módszert alkalmazták a modell globális értelmezésére és a jellemzők fontosságának becslésére (Mingyu 2021). Hasonlóképpen, Karlos és munkatársai a hallgatók végső osztályzatát jelezték előre egy online kurzuson, és SHAP értékeket használtak a jellemzők globális hozzájárulásának meghatározására (Karlos–Kostopoulos–Kotsiantis 2020). Egy szorosan kapcsolódó munkában, Smith, Chimedza és Bührmann nemcsak a változók globális fontosságát és általános hatását vizsgálták, hanem mind a SHAP-et, mind a LIME-ot felhasználták egy olyan modell előrejelzéseinek lokális magyarázatára, amelyet egy kurzuson való lemorzsolódásban veszélyeztetett hallgatók azonosítására terveztek (Smith–Chimedza–Bührmann 2021). Bár úgy tűnik, hogy a LIME algoritmust gyakrabban használják a kapcsolódó munkákban, Smith és munkatársai, valamint Molnar azt javasolják, hogy a LIME helyett érdemesebb a SHAP-et használni a lokális előrejelzések magyarázatához, mivel a SHAP stabilabbnak bizonyult, mint a LIME. Ezért ebben a munkában a SHAP módszerén keresztül mutatjuk be az egyéni előrejelzések magyarázatának folyamatát (Smith–Chimedza–Bührmann 2021; Molnar 2020).

Míg a kapcsolódó munkák a hallgatók tanulmányi átlagát és az egyes kurzusokon nyújtott teljesítményt jelzik előre, ebben a tanulmányban a kimeneti változó az egyetemi hallgatók végső tanulmányi státusza. Tehát a célunk megkülönböztetni a várhatóan diplomát szerző és a lemorzsolódás veszélyének kitett hallgatókat. Ez a munka a korábbi konferenciadolgozataink (Baranyi–Nagy–Molontay 2020; Nagy–Molontay–Szabó 2019) kibővítésének tekinthető. Korábban azt vizsgáltuk, hogyan lehet mélytanulási algoritmusokat felhasználni a globálisan értelmezhető lemorzsolódás-előrejelzéshez (Baranyi–Nagy–Molontay 2020). Ezenkívül bemutattunk egy webes alkalmazást is az értelmezhető lemorzsolódás-előrejelzéshez (Nagy–Molontay–Szabó 2019). A jelen munka fő hozzájárulása az, hogy alaposan bemutatjuk, hogyan használhatók a megmagyarázható mesterséges-intelligencia-eszközök az értelmezhető lemorzsolódás-előrejelzéshez, és ami még fontosabb, hogy személyre szabott visszajelzést adjunk a hallgatóknak, kiemelve, hogy mely készségeket kell fejleszteniük a diplomaszerezés esélyeinek növelése érdekében. Ezen kívül azt is bemutatjuk, hogy hogyan tudnak beépülni a KRÉTA országos tanulmányi rendszerébe az általunk fejlesztett módszerek és hogy a gyakorlatban hogyan tud közvetlenül hasznosulni a kutatás eredménye.

A lemorzsolódás előrejelzéséhez fejlesztett „fekete doboz” gépi tanulási modellünket (CatBoost) globálisan és lokálisan is értelmezzük. A globális értelmezés azt jelenti, hogy azonosítjuk, mely változók befolyásolták legjobban a kimenetet, és megvizsgáljuk egyes változók inte-

rakciójának hatását is a modell kimenetére (predikciójára). A fentebb idézett kapcsolódó munkák többségével ellentétben ez a munka azonban a modell globális értelmezése mellett a predikciók lokális magyarázatára is koncentrálna. Nevezetesen, a SHAP-értékek erejét felhasználva vizsgáljuk az egyes individuális előrejelzések mögött meghúzódó lehetséges okokat.

Adatok és módszertan

Adatok

A tanulmány a Budapesti Műszaki és Gazdaságtudományi Egyetem (BME) anonimizált adatain alapul. A BME-n a következő három képzési területeken kínálnak képzéseket: műszaki (mérnöki vagy informatikai), gazdaság- és társadalomtudományi, illetve természettudományi. Az adathalmaz azokat a hallgatókat tartalmazza, akik 2013 és 2017 között iratkoztak be valamelyik alapképzésre, és azóta valamilyen úton befejezték tanulmányaikat: vagy diplomát szereztek, vagy lemorzsolódtak.

Először elvégeztünk néhány adat-előkészítési lépést: Töröltük a hiányos adatokat tartalmazó sorokat, csak a hallgatók „első próbálkozását” vettük figyelembe, azaz ha egy hallgató kibukott, majd újra beiratkozott, csak az első szak eredményét vettük figyelembe. Kizártuk azokat a hallgatókat is, akiknél a lemorzsolódás oka a képzésváltás volt. Az előkészített adathalmaz összesen 8508 rekordot tartalmaz.

Adathalmazunk változói többnyire a beiratkozás előtti tanulmányi teljesítménymutatók, amelyek a felvételi pontszám komponensei, tehát amiken az egyetemi felvétel múlik (Nagy–Molontay 2021). Pontosabban a változók többsége a középiskolai osztályzatokra és az érettségi vizsgán elért eredményekre vonatkozik. Az adathalmaz dimenziójának, illetve a változók közti redundancia csökkentése érdekében, a középiskolai osztályzatok tantárgyankénti kezelése helyett a tanulmányi átlagot vizsgáljuk, ami a következő tárgyakat foglalja magában: történelem, matematika, magyar nyelv és irodalom, egy idegen nyelv és egy választott természettudományos tárgy. Az érettségi négy alaptárgyból és legalább egy választott tantárgyból áll. Az alaptantárgyak a történelem, a magyar nyelv és irodalom, egy idegen nyelv és a matematika.

Eredetileg az érettségi vizsgán elérhető pontszám 0 és 100 között lehet, azonban a diákok közép és emelt szintű érettségi vizsgát is tehetnek. Hogy elkerüljük a vizsga szintjét jelző további változók használatát (amivel a dimenziót növelnénk), emelt szintű vizsga esetén a pontszámot megszoroztuk 1,35-tel, ez a jutalmazási séma bizonyult a legnagyobb prediktív erővel rendelkezőnek a diákok végső egyetemi sikerességére vonatkozóan (Molontay–Nagy 2022). Így a 100 feletti érettségivizsga-eredmény szükségszerűen azt jelenti, hogy a tanuló emelt szintű vizsgát tett.

Az adathalmaz tartalmaz egy „Nyelvvizsga” nevű változót, amely a tanuló által szerzett nyelvvizsga-bizonyítványok mennyiségét és szintjét méri. Az eredményeket úgy kódoltuk, hogy egy B2 szintű komplex nyelvvizsga egy pontot ér, és ha a tanulónak csak „fél” nyelvvizsgája van (csak írásbeli/szövebeli vizsga), akkor az fél pontot ér. A C1 szintű nyelvvizsgák pontszáma kétszer annyi pontot ér, mint a B2 szintű vizsgáké. Ezeket a pontszámokat minden hallgatóra kiszámoljuk, és összegezzük hallgatónként.

A beiratkozás pillanatában rendelkezésre álló tanulmányi teljesítménymutatók mellett a hallgatók nemét is felhasználtuk az elemzés során. Az érettségi és a beiratkozás dátuma alapján definiáltunk egy új, „Eltelt év” nevű változót, ami azt méri, hogy hány év telt el az érettségi és a beiratkozás között.

Végezetül, a célváltozónk, a „Státusz” egy olyan bináris változó, amelynek értéke egy, ha a hallgató diplomát szerzett, és nulla, ha lemorzsolódott.

Összefoglalva, az adatkészletünk a következő változókat tartalmazza: Matematika (a matematika érettségi eredménye), Választott tárgy (a választott tantárgyból elért érettségi eredménye), Magyar (a magyar nyelvből és irodalomból elért érettségi eredménye), Történelem (érettségi eredmény történelemből), Idegen nyelv (érettségi eredmény idegen nyelvből), KTÁtlag (középiskolai tanulmányi átlag), Nyelvvizsga (nyelvvizsgapontok), Eltelt év, Férfi (1 a férfiaknál és 0 a nőknél), Műszaki (1, ha a képzési terület mérnöki vagy informatikai, 0 egyébként), Term.tud. (1, ha a képzési terület természettudományi, 0 egyébként). Ha mind a Műszaki, mind a Term.tud. változó értéke nulla, az azt jelenti, hogy a képzési terület gazdaság- és társadalomtudományi. Végül a célváltozó, a Státusz, a végső egyetemi státuszt jelöli: 1 a diplomások és 0 a lemorzsolódók esetében.

Módszertan

Az úgynevezett adatszivárgás elkerülése érdekében – a kapcsolódó munkák túlnyomó többségével ellentétben – nem véletlenszerű tanítás/teszt felosztással vagy keresztvalidálással értékeljük ki a gépi tanulási modelljeinket, hanem az adatokat a beiratkozási év mentén osztjuk fel. Nevezetesen, a modellt a 2013 és 2016 között beiratkozott hallgatói csoportokon (6398 sor) tanítjuk be, és a modell teljesítményét a 2017-ben beiratkozott hallgatói kohorszon (2110 megfigyelés) értékeljük ki. Ezzel az eljárással valósabb képet kapunk arról, hogy a modell hogyan teljesítene valós környezetben való implementálásakor. A tanító és a tesztalacson a diplomát szerzettek aránya 65%, illetve 63%.

Ahogy korábban is említettük, számos önmagában is megmagyarázható modell létezik, mint például a logisztikus regresszió és a kis mélységű döntési fák, azonban ezek általában gyenge teljesítményt nyújtanak az összetett, nehezen értelmezhető gépi tanulási modellekhez képest (pl. gradient boosting trees és neurális hálózatok).

A modell-agnosztikus IML-eszközöknek köszönhetően azonban a teljesítmény és az értelmezhetőség közötti kompromisszum már nem jelent többé problémát, hiszen a post hoc IML-eszközök lehetővé teszik, hogy a fekete doboz modelleket értelmezhetővé és átláthatóvá tegyük.

Számos gépi tanulási algoritmust teszteltünk (teljesítményüket az 1. táblázat részletezi), és az előrejelzések elkészítéséhez a legjobban teljesítő modellt, a CatBoost algoritmust használjuk, amely jelenleg az egyik legkorszerűbb modellnek számít a táblázatos adathalmazok esetében (Prokhorenkova et al. 2018). A CatBoost modell értelmezéséhez olyan modern technikákat használunk, mint a permutációs fontosság (permutation importance) (Fisher–Rudin–Dominici 2019), kétdimenziós parciális függőségi ábrák (partial dependence plot) (Greenwell–Boehmke–McCarthy 2018) és SHAP (SHapley Additive exPlanations) értékek (Lundberg–Lee 2017), amelyek a Shapley-érték játékelméleti koncepcióján alapulnak.

A permutációs fontosság (PI) egy egyszerű módszer a változók előrejelzésre gyakorolt hatásának mérésére (Fisher–Rudin–Dominici 2019). A modell betanítása után a következőképpen számítjuk ki a változó fontosságát: megkeverjük (permutáljuk) a vizsgálandó változó értékeit, az összes többi változót és a célváltozót változatlanul hagyva, és megmérjük, hogy a modell teljesítménye mennyire romlik a megkevert adathalmazon. Minél jobban csökken a teljesítmény a keverés miatt, annál fontosabb a jellemző. A módszer véletlenszerűségének kiiktatásához a változókhoz rendelt fontossági mértéket 100-szor számoljuk újra, és ebből átlagot vonunk.

A kétdimenziós parciális függőségi diagram (PDP) azt mutatja, hogy két jellemző milyen marginális (parciális) hatással van a modell előrejelzésére. A PDP elkészítése úgy történik, hogy ábrázoljuk, hogyan változik a már betanított modell előrejelzése, ha megváltoztatjuk a kiválasztott változók értékeit (Greenwell–Boehmke–McCarthy 2018).

Míg a PI és a PDP segít globálisan értelmezni az alapul szolgáló „fekete doboz” gépi tanulási modellt, a SHAP-értékek ezen kívül arra is felhasználhatók, hogy lebontsák, hogyan működik a modell egy-egy egyéni előrejelzés esetében. A SHAP-érték egy játékelméleti fogalmon, a Shapley-értéken alapul, és a változóknak az előrejelzésre gyakorolt hatását (hozzájárulását) mutatja. Pontosabban a jellemzők hozzájárulásának meghatározásánál azt hasonlítja össze, hogy mennyivel változik a modell predikciója a változó értékének ismeretében a modell átlagos előrejelzéséhez képest, amelyet alapértéknek nevezünk. A SHAP-érték tehát egy lokális előrejelzést úgy magyaráz meg, hogy megmutatja, hogyan és milyen mértékben mozdítják el a hallgató változóinak értékei a modell előrejelzését az alapértékhez viszonyítva (Lundberg–Lee 2017).

Megjegyzendő, hogy óvatosan kell bánni a megmagyarázható mesterségesintelligencia-eszközök használatával. Például mind a PI, mind a PDP feltételezi, hogy a

jellemzők statisztikailag függetlenek, ezért egymással erősen korreláló változók esetén félrevezető értelmezésekhez vezethetnek (Molnar et al. 2020). Továbbá Slack és munkatársai megmutatták, hogy a LIME és SHAP segítségével szándékosan félrevezető értelmezések hozhatók létre, amelyek elrejtetik a modell torzításait (Slack et al. 2020). Végül Kumar és munkatársai rávilágítottak a Shapley-érték-alapú jellemző fontossági mértékek használatának problémáira/gyengeségeire, és megmutatták, hogy a megoldás további komplexitást indukálna (Kumar 2020).

Eredmények

Ebben a fejezetben a BME adatain keresztül bemutatjuk, hogy a megmagyarázható mesterségesintelligencia-eszközök hogyan segíthetnek feltárni a lemorzsolódás előrejelzését leginkább befolyásoló tényezők hatását, és hogyan lehet ezen eszközök kimenetét személyre szabott beavatkozásra és visszajelzés nyújtására használni.

Első körben a gépi tanulási modellünk globális értelmezésére összpontosítunk a változók fontosságának meghatározásával. A SHAP-értékek segítségével elemezzük, hogy a jellemzők általában hogyan befolyásolják a diplomaszerezés valószínűségét, továbbá a változók interakcióját is vizsgáljuk 2D parciális függőségi ábrák segítségével. Ezután a következő alfejezetben bemutatjuk, hogy a SHAP-értékek felhasználásával hogyan lehet egyéni visszajelzéseket adni.

Modellválasztás

Számos gépi tanulási modellt teszteltünk, többek között az XGBoostot, Lineáris diszkriminanciaanalízist (Linear discriminant analysis), a CatBoostot, az AdaBoostot, az NGBoostot stb., és azt találtuk, hogy az adathalmazunkon a CatBoost nyújtja a legjobb teljesítményt (lásd az 1. táblázatot), ezért a továbbiakban a CatBoost modell optimalizálásával és alkalmazásával foglalkozunk.

A tesztadathalmazon az Optuna (Akiba et al. 2019) segítségével optimalizált CatBoost 0.847-es átlagos pontosságot (average precision) és 0.774-es AUC értéket ért el. Az elért teljesítmény összhangban van a kapcsolódó munkákkal, amelyek nagy méretű heterogén adathalmazon végeznek lemorzsolódás-előrejelzést, például Behr és munkatársai 0.77-es AUC-t értek el (Behr et al. 2020), továbbá Lee és Chung legjobban teljesítő modelljének átlagos pontossága 0.898 (Lee-Chung 2019).

Globális értelmezés

Első körben a CatBoost beépített algoritmusával, valamint a permutációs fontosság módszerével számoltuk ki a változók fontosságát (az eredmények a 2. táblázatban találhatóak). Mindkét fontossági metrika szerint a legfontosabb jellemzők a középiskolai tanulmányi átlag,

1. táblázat | A gépi tanulási modellek teljesítménye (AUC és átlagos pontosság) a tesztadathalmazon

Modell	AUC-érték	Átlagos pontosság
CatBoost	0,774	0,847 (1)
NGBoost	0,767	0,838 (2)
Explainable Boosting Modell	0,760	0,833 (3)
Logisztikus regresszió	0,734	0,799 (5)
XGBoost	0,732	0,813 (4)
Gradient Boosting Modell	0,728	0,714 (7)
Linear Discriminant Analysis	0,722	0,710 (10)
AdaBoost	0,717	0,719 (9)
Light Gradient Boosting Machine	0,708	0,714 (6)
Quadratic Discriminant Analysis	0,707	0,711 (8)

A modellek az AUC-értékük szerint vannak rendezve, az átlagos pontosság szerinti rangsoruk pedig zárójelben van feltüntetve (Forrás: saját forrás)

valamint a matematika és a választott tantárgyból elért érettségi eredmények. Megjegyzendő, hogy a változók közötti magas korreláció eltorzíthatja a változók fontosságának mértékét (Molnar et al. 2020), valószínűleg ez az oka annak, hogy a humán tantárgyak (idegen nyelv, történelem, magyar nyelv és irodalom) fontossága különösen a PI szerint alacsony, hiszen ezek között a változók között a Pearson-korreláció 0.5 feletti. Továbbá ezek a változók erősen korrelálnak a középiskolai tanulmányi átlaggal is. Például a magyar nyelv és irodalomból tett érettségi eredmény és a tanulmányi átlag között 0.64 a korreláció, ezzel szemben a matematika és a választott tárgy érettségi eredménnyel vett korrelációja a tanulmányi átlagnak csupán 0.39 és 0.31.

2. táblázat | A jellemzők fontossága a CatBoost és a PI szerint

Változó	Változó fontosság CatBoost szerint	Permutációs fontosság
KTÁtlag	18,321	0,055 (1)
Matematika	18,185	0,044 (2)
Eltelt évek	17,766	0,018 (6)
Választott tárgy	11,618	0,026 (3)
Férfi	8,859	0,019 (5)
Műszaki	7,323	0,023 (4)
Idegen nyelv	5,462	0,003 (8)
Történelem	4,301	0,003 (9)
Magyar	3,353	0,001 (11)
Nyelvvizsga	2,893	0,003 (10)
Term.tud.	1,920	0,004 (7)

A permutációs fontosság a tesztadathalmazon lett becsülve az AUC érték alapján. A permutációkat 100-szor ismételtük meg, és a táblázatban az átlagos fontosság látható. A változók a CatBoost fontosság szerint vannak rendezve, a PI szerinti rangsoruk zárójelben szerepel (Forrás: saját forrás)

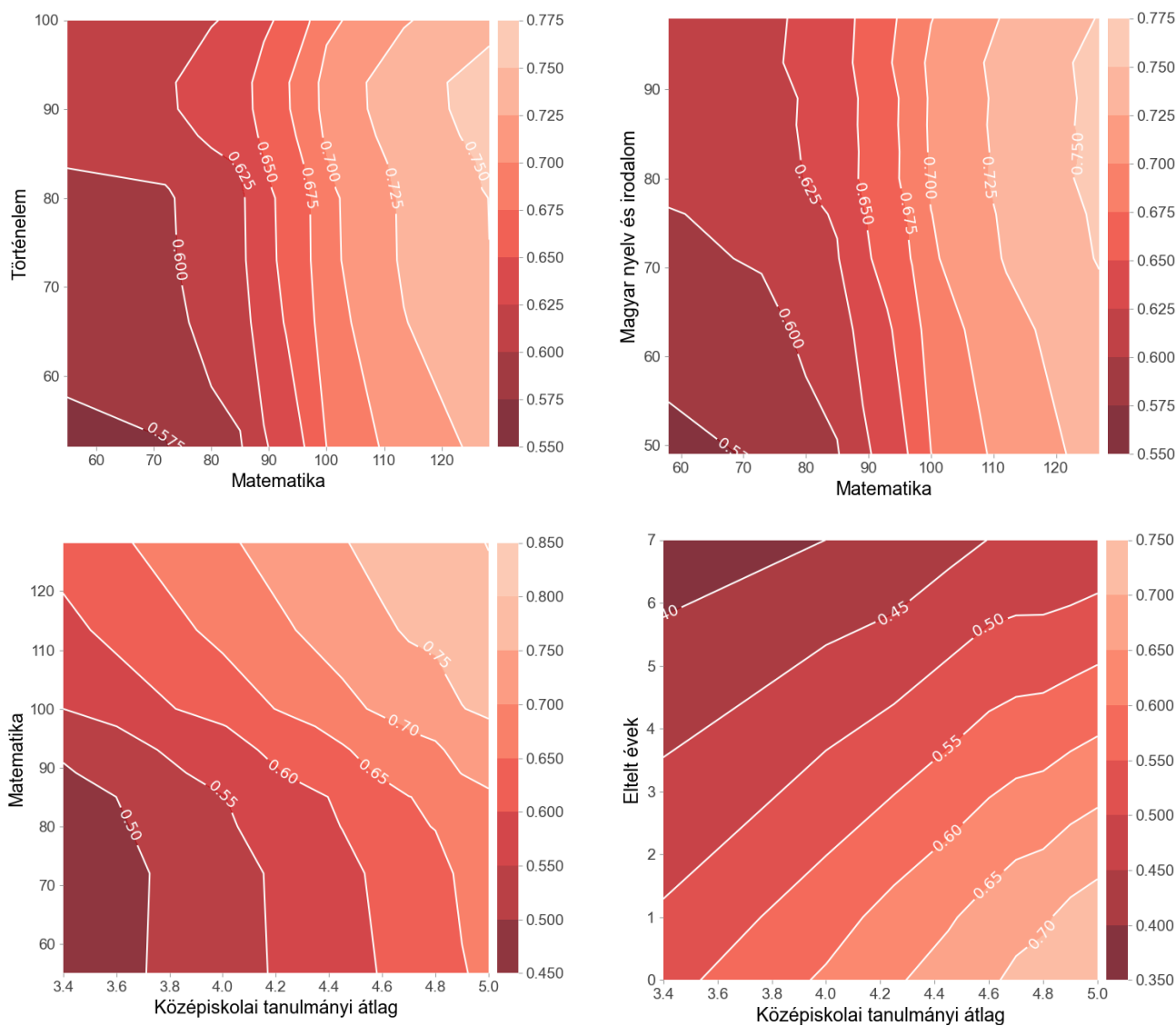
A változók előrejelzésre gyakorolt együttes hatásának megértéséhez a kétdimenziós parciális függőségi ábrákhoz fordulunk. Az 1. ábra bal felső diagramja a matematika és történelem érettségi pontszámok különböző kombinációihoz tartozó előrejelzéseket mutatja. Az ábra azt sugallja, hogy minél jobb a matematika érettségi eredmény, annál nagyobb a CatBoost modell kimenete, azaz a végzés jóslat valószínűsége. Továbbá a legmagasabb végzési valószínűség akkor érhető el, ha a történelem érettségien elért pontszám is viszonylag magas. Azonban, ha valaki kiemelkedő eredményt ér el emelt szintű matematika érettségien, és a történelem érettségien elért pontszáma „túl” magas, akkor a diplomaszerezés előre jelzett valószínűsége kissé csökken.

Hasonlóképpen az 1. ábra jobb felső kontúrdiagramja a magyar nyelv és irodalom és a matematika érettségi

pontszámok közötti interakciót mutatja. Nem meglepő, hogy egy műszaki egyetemen a matematikai készségek erősen befolyásolják a modell előrejelzését, azonban a modell akkor prediktálja a legmagasabb végzési valószínűséget, ha a magyar nyelv érettségi pontszáma is magas.

Az 1. ábra bal alsó diagramja a középiskolai tanulmányi átlag és a matematika érettségi eredmény interakcióját mutatja. Az ábra alapján világos, hogy ezek a változók egyformán fontosak, és a diákok nagyobb eséllyel szereznek diplomát, ha magas pontszámot érnek el matematikából és kitűnő osztályzatokat szereznek középiskolai tanulmányaik során.

A jobb alsó ábra arra utal, hogy ha a végzős középiskolások nem kezdik meg egyetemi tanulmányaikat közvetlenül az érettségi után, akkor a kihagyott évek számának növelésével egyre csökken a diplomaszerezés esélye. To-



1. ábra

A matematika interakciós kétdimenziós parciális függőségi diagramja a magyar nyelv és irodalommal és a történelemmel (felső ábrák) és a középiskolai tanulmányi átlag interakciója a matematikával és az eltelt évekkel (alsó ábrák)

Forrás: saját szerkesztés.

vábbá, az átlós kontúrvonalak azt jelzik, hogy a tanulmányi átlag és az „Eltelt évek” összemérhetően nagy hatással vannak a modell kimenetére.

A 2. ábra a változók hatását szemlélteti a modell predikciójára vonatkozóan a SHAP értékek szerint. Az ábra azt mutatja, hogy a legfontosabb változó a középiskolai tanulmányi átlag, ami összhangban van a permutációs fontosság eredményeivel (2. táblázat). Korábbi munkánkban is azt tapasztaltuk, hogy az egyetemi felvételi pontszám összetevői közül a középiskolai tanulmányi átlagnak van a legnagyobb prediktív ereje a végső tanulmányi sikerre vonatkozóan (Nagy–Molontay 2021). Az ábrán jól látható, hogy minél magasabb a tanulmányi átlag, annál valószínűbb a diplomaszerezés, hiszen – ahogy a színek mutatják – ha magas a változó értéke, az magasabbra tolja az előrejelzést, az alacsony értékek pedig alacsonyabbra húzzák a modell előrejelzését.

Mivel a vizsgált egyetem egy műszaki egyetem, nem annyira meglepő, hogy a matematikai készség nagymértékben képes előre jelezni a végső tanulmányi sikert. Jól ismert globális probléma, hogy a STEM-képzéseken a matematika készségek hiánya a lemorzsolódás és a képzésidőn túli befejezés fő oka (Baranyi–Molontay 2021). Hasonló összefüggés van a választott tantárgy és diplomaszerezés valószínűsége között is. Ez valószínűleg annak köszönhető, hogy a felvételi rendszerben a választott érettségi tárgynak a legtöbb esetben kapcsolódnia kell a szakterülethez (az elfogadható tárgyak körét az alapképzések határozzák meg).

Megfigyelhetjük, hogy a 2. ábra szerint az a tény, hogy valaki férfi, negatív hatással van a modell kimenetére. Ez

szintén összhangban van korábbi munkánk eredményével (Nagy–Molontay 2021), miszerint a nők nagyobb valószínűséggel szereznek diplomát a BME-n, mint a férfiak.

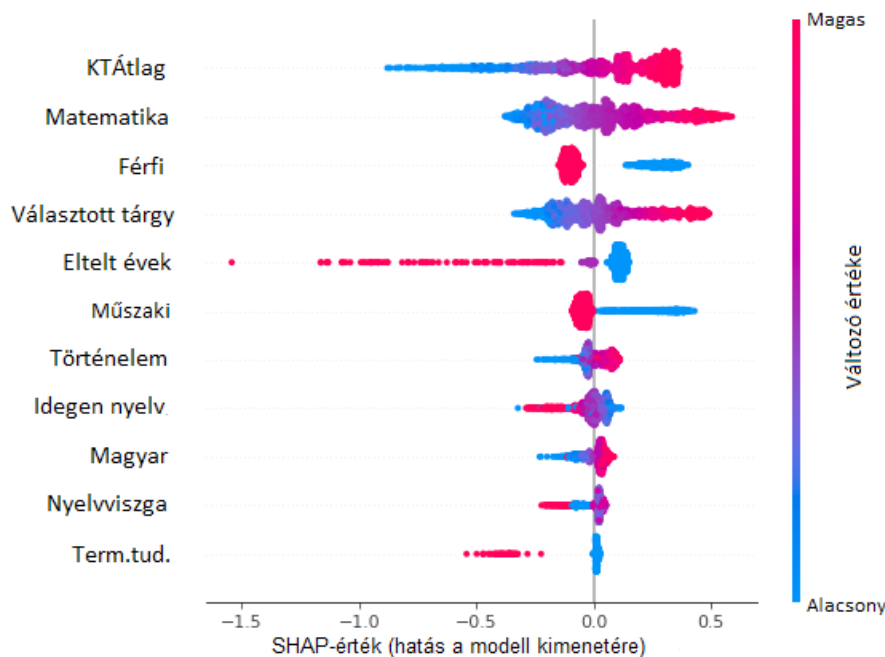
Meglehetősen meglepő módon a SHAP-értékek (lásd 2. ábra) azt mutatják, hogy az idegen nyelvi érettségi vizsga eredménye szignifikánsan negatív hatással van a diplomaszerezés valószínűségére. Ez feltehetőleg annak a következménye, hogy a humán tantárgyak között erős a korreláció, azonban ennek az okának kiderítése további részletesebb vizsgálatot igényel.

A magyar és a történelem érettségi vizsgák esetében magasabb pontszám magasabb diplomaszerezési valószínűséggel jár, azonban ezeknek a változóknak a hatása a modell kimenetére kisebb a matematika és a választott tantárgyhoz képest.

Ha a szakterület műszaki vagy természettudományi, akkor az negatívan befolyásolja a diplomaszerezés valószínűségét. Ha azonban mindkét mutató értéke nulla – vagyis a képzési terület gazdaság- és társadalomtudományi –, akkor az pozitív hatással van a végzés valószínűségére. Ennek az az oka, hogy a végzési arány szakonként változó, és a BME-n a gazdaság- és társadalomtudományi képzési terület esetén a legmagasabb.

Egyéni előrejelzések megmagyarázása

Ebben a fejezetben bemutatjuk, hogy a „fekete doboz” modellek lokális értelmezése hogyan segíthet személyre szabott visszajelzést adni az egyetemi hallgatóknak az erősségeikről és gyengeségeikről. Más szóval, bemutat-



2. ábra

A változók hatását összegző grafikon. Minden hallgatónak megfelelő egy-egy pont minden sorban. A pontok színe a változó értékét jelöli, az x-tengely pedig a modell előrejelzésére gyakorolt hatást mutatja. A jellemzők a fontosságuk (a modell kimenetére gyakorolt átlagos abszolút hatás) szerint vannak rendezve

Forrás: saját szerkesztés.

jük, hogy a SHAP-értékek hogyan használhatóak az egyéni előrejelzések magyarázatára.

A 3. ábra három hallgatón mutatja be az egyéni előrejelzések magyarázatát SHAP-értékeken keresztül. Az előrejelzést negatívan befolyásoló tényezők késsel, a diplomaszerzés valószínűségét növelő tényezők pedig pirossal vannak jelölve. A gépi tanulási modell kimenete (végzés valószínűsége) félkövér betűkkel van írva.

Az első hallgató esetében a modell kimenete 0.27, ami azt jelenti, hogy az adott tanuló esetében magas a lemorzsolódás kockázata. A SHAP-értékek azonban azt is elárulják, hogy miért van az adott hallgató kitéve a lemorzsolódás veszélyének. Nincs egyetlen olyan faktor sem, ami magasabbra tolná az előrejelzést, de az a tény, hogy a középiskola és az egyetem között 4 év telt el, valamint hogy a matematika érettségi eredmény és a tanulmányi átlag viszonylag alacsony, jelentősen csökkenti a modell előrejelzését. Az alacsony tanulmányi átlag és a matematikából és a választott tantárgyból elért alacsony érettségi eredmények arra utalnak, hogy a diáknak mind az általános, mind a programspecifikus ismeretekben fejlődnie kellene, ráadásul az a tény, hogy 4 év kihagyás után kezdte meg az egyetemet, arra utal, hogy a felzárkóztató kurzusok rendkívül hasznosak lennének ennek a lemorzsolódásban különösen veszélyeztetett diáknak.

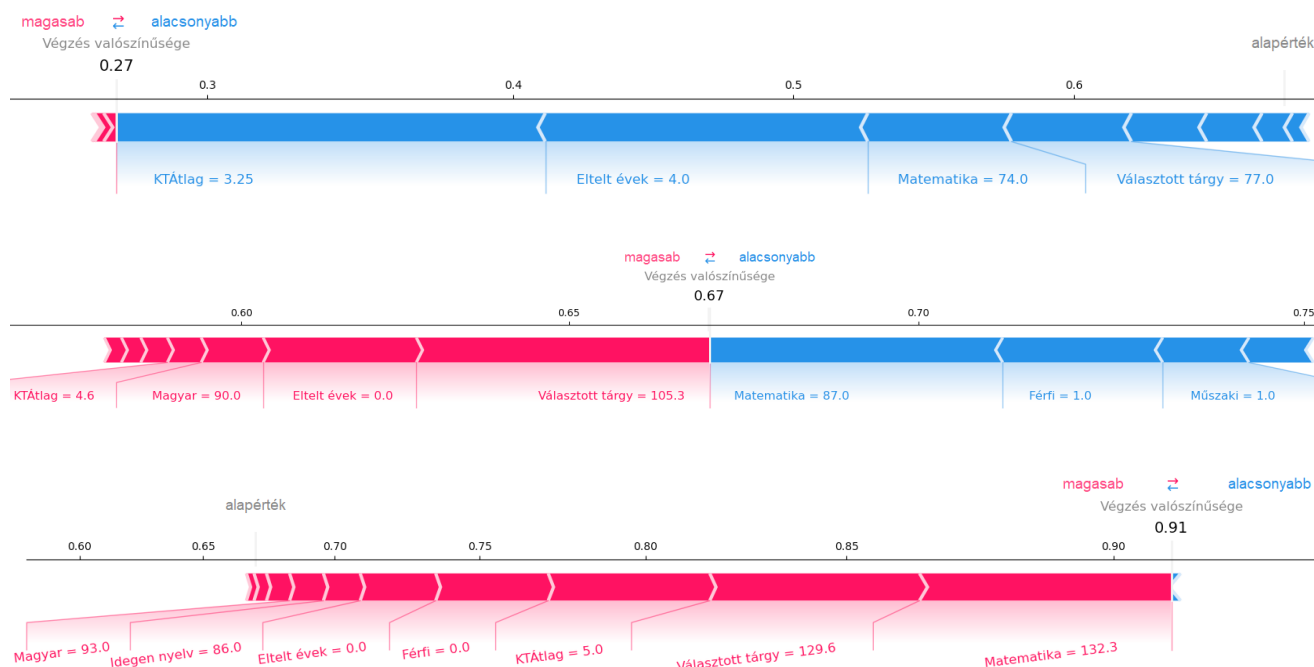
A 3. ábra második példája egy közepes diákot mutat be, akinél az előrejelzés megegyezik az alapértékkel (átlagos előrejelzés). A diák érettségi pontszáma a választott tantárgyból magas, ráadásul a középiskola után azonnal egyetemen folytatta tanulmányait, ami nagy-

mértékben pozitívan befolyásolja az előrejelzést. Másrészt a viszonylag alacsony matematika és történelem pontszámok lejjebb tolják az előrejelzést, és kioltják a pozitív tényezők hatását. Továbbá, mivel a férfi hallgatók kisebb valószínűséggel szereznek diplomát a BME-n, mint a nők, a hallgató neme is negatívan befolyásolja az előrejelzést. Mivel a matematika pontszámának van a legjelentősebb negatív hatása, személyre szabott beavatkozási tervként matematikai felzárkóztató kurzust kell ajánlani a diák számára.

Végül a 3. ábra utolsó példája egy tehetséges diákot mutat be, akinek a kitűnő tanulmányi átlaga és a kiváló érettségi eredménye matematikából és a választott tantárgyból jelentős pozitív hozzájárulással bír az előrejelzéshez. Emellett a már említett tény, hogy a nők nagyobb valószínűséggel szereznek diplomát a BME-n, még magasabbra tolja a végzés valószínűségét.

Gyakorlati hasznosulás és megvalósulás

A Kooperatív Doktori Program keretében az eKRÉTA Informatikai Zrt. munkatársaival közösen fejlesztünk egy mesterségesintelligencia-alapú modult, ami a műegyetemi adatokon végzett előzménykutatásokon alapszik, azonban célunk, hogy az összes magyar felsőoktatási intézményre elérhető legyen. A KRÉTA Tanulmányi Rendszerébe beépülő modul elsődleges célja, hogy a felsőbb éves középiskolás diákok számára megmutassa, hogy milyen készségeket kell fejleszteniük ahhoz, hogy gördülékenyebb legyen a diplomáig vezető út. A modul



3. ábra

Három diák egyéni előrejelzéseinek magyarázata a SHAP-értékek segítségével. Az alapérték (végzés átlagos valószínűsége a modell szerint) 0.67. A modell predikcióját az alapértéktől magasabbra toló faktorok piros színnel, az előrejelzést alacsonyabbra húzó tényezők pedig késsel vannak jelölve. A sávok hossza megfelel a változók hozzájárulásának mértékének

Forrás: saját szerkesztés

szíve egy komplex fekete doboz algoritmus, aminek kimenetét megmagyarázható mesterséges intelligencia (XAI) módszerek segítségével tesszük érthetővé a közönség számára. A modul inputja a középiskolai osztályzatok, az érettségi eredmények (valós vagy várt), valamint egy választott egyetemi képzés. Ezek függvényében, hasonlóan a 3. ábrán bemutatott egyéni előrejelzések megmagyarázásához, a modul megmutatja, hogy mely tantárgyak eredményén kellene javítani ahhoz, hogy sikeresen vegye az akadályokat. A modul elsődleges célközönsége a középiskolai diákok, de hasznos lehet a felsőoktatás elsőéves hallgatói számára is, például, ha a modul szerint a matematikai készségeken kell javítani (hasonlóan a 3. ábra második példájához), akkor a hallgatónak ajánlott beiratkozni a felzárkóztató matematika kurzusra. Az alkalmazás előzetes verzióját már korábban bemutattuk az egyetem vezetőségének is, akik azt kifejezetten hasznosnak és könnyen érthetőnek találták. Idén ősszel kerül sor a KRÉTA-ba beépített modul első verziójának tesztelésére a felhasználók bizonyos körében, ami után rövidesen elérhető lesz a modul a teljes publikum számára is.

Bár egy mesterségesintelligencia-alapú döntéstámogató rendszer hatékony eszköz lehet a felsőoktatásban a végzési arány növelésére, fontos megjegyezni, hogy a prediktív analitika alkalmazása számos etikai dilemmát is felvet. A mi esetünkben a legfőbb etikai kérdés az algoritmus kimeneti eredményének, azaz a végzés valószínűségének használatában és kommunikációjában rejlik. Nyilvánvaló, hogy nem tájékoztathatjuk a diákokat nyersen a megjósolt végső tanulmányi eredményükről, mivel ha valakinek a tanulmányai kezdetén közöljük, hogy 0,6-os valószínűséggel le fog morzsolódnani, az rendkívül etikátlan lenne, és visszafelé süllhet el, önbeteljesítő jóslatokhoz vezethet, és ezáltal csak növelné a lemorzsolódási arányt. Emiatt és más technikai okok miatt az eKRÉTA-val fejlesztett modulban, a modell kimenete nem a végzés valószínűsége, hanem egy általunk definiált sikerességi mutató, ami a kreditindexen alapul. Továbbá, nem nyersen közöljük a hallgató becsült eredményét, hanem azt, hogy a szakra beiratkozott hallgatókhoz képest hányadik decilisben/percentilisben helyezkedik el az illető. Emellett, a modul kimenetében a predikció magyarázatára összpontosítunk, amely jelzi, hogy milyen készségeket kell fejlesztenie az egyénnek.

Összefoglalás és konklúzió

Ebben a munkában a lemorzsolódásban veszélyeztetett hallgatók azonosításán keresztül bemutattuk, hogy hogyan lehet a magyarázható mesterséges intelligencia eszközeivel a „fekete doboz” algoritmusok előrejelzését átláthatóvá és érthetővé tenni.

A globális értelmezések segítségével rávilágítottunk azokra a tényezőkre, amelyek a legnagyobb hatással vannak a tanulmányi teljesítményre. A permutációs fontosság (2. táblázat) és a SHAP fontosság (2. ábra) is azt

mutatta, hogy a végső tanulmányi státuszra a legnagyobb előre jelző erővel rendelkező változó a középiskolai tanulmányi átlag, ami az általános tudást méri, mivel figyelembe veszi a történelem, a matematika, a magyar nyelv és irodalom, egy idegen nyelv és egy természettudományos tárgy osztályzatait. Az a tény azonban, hogy a matematika és a választott tantárgy is a legfontosabb változók között van, arra utal, hogy a programspecifikus tudás sem elhanyagolható, és kiegészíti az általános tudást. Azt is megállapítottuk, hogy a diákok nagyobb valószínűséggel morzsolódnak le, ha nem közvetlenül a középiskola elvégzése után kezdik meg egyetemi tanulmányaikat. A parciális függőségi diagram segítségével megmutattuk, hogy a humán tantárgyaknak is van inkrementális prediktív erejük, annak ellenére, hogy ez az elemzés egy műszaki egyetem adataira épül.

Eljárásunk újdonsága, hogy nemcsak a végzés valószínűségét becsüljük meg, hanem interpretálhatóvá tesszük a modellt, és az egyéni előrejelzéseket is megmagyarázzuk. Ez teszi ezt a keretrendszert több különböző területen is alkalmazhatóvá. Például segíthet a középiskolásoknak a megfelelő szak kiválasztásában az egyetemen, és pályaeorientációs tanácsadást nyújthat azáltal, hogy a középiskolai teljesítménymutatók alapján megbecsüli, hogy melyik szakon lehetnek a legsikeresebbek. Ezenkívül a predikciók lokális magyarázatával a bemutatott keretrendszer abban is segítheti a diákokat, hogy azonosítsák azokat a készségeket, amelyeket fejleszteni kell ahhoz, hogy sikeresek legyenek az egyetemi tanulmányaikban. Továbbá a bemutatott módszertan a felsőoktatási döntéshozók számára is hasznos lehet, hogy megtalálják a megfelelő beavatkozási pontokat a veszélyeztetett diákok számára, például személyre szabott korrepetálás vagy felzárkóztató kurzusok ajánlásával.

A tanulmány egyik limitációja, hogy kizárólag a Műegyetem adataira épül, ahol a legtöbb hallgató műszaki képzésben vesz részt, ezért érdekes lenne más felsőoktatási intézmények adatain is elvégezni az elemzéseket. Jövőbeni kutatási tervünk, hogy a legtöbb magyarországi felsőoktatási intézmény adataira alapozva világos és átfogó képet adjunk az egyetemi sikerességhez leginkább hozzájáruló tényezőkről. Emellett fontos megjegyezni, hogy míg a hallgatói lemorzsolódást számos tényező befolyásolja, addig mi csak néhány, főként a középiskolai teljesítményhez kapcsolódó változókkal rendelkezünk. Más változókat is összefüggésbe hoztak a hallgatói sikerrel, mint például a társadalmi-gazdasági státusz (Zwick–Himelfarb 2011; Freitas et al. 2020), a kollégiumi lét (Zeleny–Molontay–Szabó 2021) és különböző pszichológiai tényezők (Séllei–Stumphauer–Molontay 2021).

Köszönetnyilvánítás

Hálás köszönet Molontay Rolandnak a cikk megírása során nyújtott segítségéért és a Kooperatív Doktori Program keretein belül végzett kutatás-fejlesztésben való konzulensi

munkájáért. Külön köszönet Szabó Mihálynak a kézirat alapos átnézéséért és a hasznos megjegyzéséért.

A kutatás az Európai Unió támogatásával valósult meg, az RRF-2.3.1-21-2022-00004 azonosítójú, Mesterséges Intelligencia Nemzeti Laboratórium projekt keretében.

A C1002329 számú projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a KDP-2020 pályázati program finanszírozásában valósult meg.



Irodalomjegyzék

- Adadi, A., & Berrada, M. (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, Vol. 66. pp. 52138–52160.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019) Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*
- Altmann, A., Tološi, L., Sander, O., & Lengauer, T. (2010) Permutation importance: a corrected feature importance measure. *Bioinformatics*, Vol. 26. Issue 10. pp. 1340–1347.
- Alyahyan, E., & Dűstegör, D. (2020) Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, Vol. 17. pp. 1–21.
- Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016) Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, Vol. 20. Issue 2. pp. 13–29.
- Baranyi, M., & Molontay, R. (2021) Comparing the effectiveness of two remedial mathematics courses using modern regression discontinuity techniques. *Interactive Learning Environments*, Vol. 29. pp. 247–269.
- Baranyi, M., Nagy, M., & Molontay, R. (2020) Interpretable Deep Learning for University Dropout Prediction. *Proceedings of the 21st Annual Conference on Information Technology Education*, pp. 13–19.
- Behr, A., Giese, M., Tegum, K. H., Theune, K. (2020) Early prediction of university dropouts—a random forest approach. *Jahrbücher für Nationalökonomie und Statistik*, Vol. 240. Issue 6. pp. 743–789.
- Cano, A., & Leonard, J. D. (2019) Interpretable multiview early warning system adapted to underrepresented student populations. *IEEE Transactions on Learning Technologies*, Vol. 12. No. 2. pp. 198–211.
- Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., & Raes, A. (2020) Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. *Decision Support Systems*, Vol. 135. 113325. <https://doi.org/10.1016/j.dss.2020.113325>
- Daniel, B. K. (2017) Overview of big data and analytics in higher education. In: *Big data and learning analytics in higher education*. pp. 1–4. Springer
- Du, M., Liu, N., & Hu, X. (2019) Techniques for interpretable machine learning. *Communications of the ACM*, Vol. 63. No. 1. pp. 68–77.
- Dutt, A., Ismail, M. A., & Herawan, T. (2017) A systematic review on educational data mining. *IEEE Access*, Vol. 5. pp. 15991–16005.
- Fisher, A., Rudin, C., & Dominici, F. (2019) All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.*, Vol. 20. No. 177. pp. 1–81.
- Freitas, F. A., Vasconcelos, F. F., Peixoto, S. A., Hassan, M. M., Dewan, M., Albuquerque, V. H., & Rebouças Filho, P. P. (2020) IoT System for School Dropout Prediction Using Machine Learning Techniques Based on Socioeconomic Data. *Electronics*, Vol. 9. No. 10. 1613.
- Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018) A simple and effective model-based variable importance measure. *arXiv preprint arXiv:1805.04755*.
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019) XAI—Explainable artificial intelligence. *Science Robotics*, Vol. 4. No. 37. <https://doi.org/10.1126/scirobotics.aay7120>
- Helal, S., Li, J., Liu, L., Ebrahimic, E., Dawson, S., & Murray, D. J. (2019) Identifying key factors of student academic performance by subgroup discovery. *International Journal of Data Science and Analytics*, Vol. 7. pp. 227–245.
- Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019) A systematic review of deep learning approaches to educational data mining. *Complexity*, Vol. 2019. <https://doi.org/10.1155/2019/1306039>
- Karimi, A.-H., Barthe, G., Balle, B., & Valera, I. (2020) Model-agnostic counterfactual explanations for consequential decisions. *International Conference on Artificial Intelligence and Statistics*. pp. 895–905.
- Karlos, S., Kostopoulos, G., & Kotsiantis, S. (2020) Predicting and interpreting students’ grades in distance higher education through a semi-regression method. *Applied Sciences*, Vol. 10. No. 23. 8413.
- Kumar, I. E., Venkatasubramanian, S., Scheidegger, C., & Friedler, S. (2020) Problems with Shapley-value-based explanations as feature importance measures. *International Conference on Machine Learning*, Vol. 119. pp. 5491–5500.
- Latif, A., Choudhary, A. I., & Hammayun, A. A. (2015) Economic effects of student dropouts: A comparative study. *Journal of Global Economics*. Vol. 3. No. 2. pp. 1–4. <https://doi.org/10.4172/2375-4389.1000137>
- Lee, S., & Chung, J. Y. (2019) The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences*, Vol. 9. No. 15. 3093.
- Lester, J., Klein, C., Rangwala, H., & Johri, A. (2017) *Learning Analytics in Higher Education: ASHE Higher Education Report*, Vol. 43. No. 5.
- Looveren, A. V., & Klaise, J. (2021) Interpretable counterfactual explanations guided by prototypes. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 650–665.
- Lundberg, S. M., & Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. In: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (eds): *Advances in Neural Information Processing Systems 30* pp. 4765–4774. Curran Associates, Inc.
- Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y., Mousa Fardoun, H., & Ventura, S. (2016) Early dropout prediction using data mining: a case study with high school students. *Expert Systems*, Vol. 33. No. 1. pp. 107–124.

- Mingyu, Z., Sutong, W., Yanzhang, W., & Dujuan, W. (2021) An interpretable prediction method for university student academic crisis warning. *Complex & Intelligent Systems*, Vol. 8. pp. 323–336.
- Molnar, C. (2020) *Interpretable Machine Learning*. Lulu.com.
- Molnar, C., König, G., Herbringer, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., ... Bischl, B. (2020) General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. arXiv preprint arXiv:2007.04131.
- Molontay, R., & Nagy, M. (2022) How to improve the predictive validity of a composite admission score? A case study from Hungary. *Assessment & Evaluation in Higher Education*, pp. 1–19.
- Mothilal, R. K., Sharma, A., & Tan, C. (2020) Explaining machine learning classifiers through diverse counterfactual explanations. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 607–617.
- Nagrech, S., Dillon, J. Z., & Chawla, N. V. (2017) MOOC Dropout Prediction: Lessons Learned from Making Pipelines Interpretable. *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 351–359.
- Nagy, M., & Molontay, R. (2021) Comprehensive analysis of the predictive validity of the university entrance score in Hungary. *Assessment & Evaluation in Higher Education*, Vol. 46. No. 8. pp. 1235–1253.
- Nagy, M., Molontay, R., & Szabó, M. (2019) A web application for predicting academic performance and identifying the contributing factors. *47th Annual Conference of SEFI*, pp. 1794–1806.
- Niculescu-Mizil, A., & Caruana, R. (2005) Predicting good probabilities with supervised learning. *Proceedings of the 22nd international conference on Machine learning*, pp. 625–632.
- Platt, J. (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, Vol. 10. pp. 61–74.
- Powell, W. W., & Snellman, K. (2004) The knowledge economy. *The Annual Review of Sociology*, Vol. 30. pp. 199–220.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018) CatBoost: unbiased boosting with categorical features. *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS'18)*, pp. 1–11.
- Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020) A survey on predictive models of learning analytics. *Procedia Computer Science*, Vol. 167. pp. 37–46.
- Rastrollo-Guerrero, J. L., Gomez-Pulido, J. A., & Durán-Domínguez, A. (2020) Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, Vol. 10. No. 3. 1042.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016) "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 13–17, 2016. pp. 1135–1144.
- Rovira, S., Puertas, E., & Igual, L. (2017) Data-driven system to predict academic grades and dropout. *PLoS ONE*, Vol. 12. No. 2. e0171207.
- Sargsyan, A., Karapetyan, A., Woon, W. L., & Alshamsi, A. (2020) Explainable AI as a Social Microscope: A Case Study on Academic Performance. *International Conference on Machine Learning, Optimization, and Data Science*, pp. 257–268.
- Séleli, B., Stumphauer, N., & Molontay, R. (2021) Traits versus Grades—The Incremental Predictive Power of Positive Psychological Factors over Pre-Enrollment Achievement Measures on Academic Performance. *Applied Sciences*, Vol. 11. Vol. 4. 1744.
- Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020) Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 180–186.
- Smith, B. I., Chimedza, C., & Bührmann, J. H. (2021) Individualized help for at-risk students using model-agnostic and counterfactual explanations. *Education and Information Technologies*, Vol. 27. pp. 1539–1558.
- Varga, E. B., & Sátán, Á. (2021) Detecting at-risk students on Computer Science bachelor programs based on pre-enrollment characteristics. *Hungarian Educational Research Journal*, Vol. 11. No. 3. pp. 297–310.
- Vultureanu-Albiși, A., & Bădică, C. (2021) Improving Students' Performance by Interpretable Explanations using Ensemble Tree-Based Approaches. *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 215–220.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019) Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, Vol. 16. pp. 1–27.
- Zeleny, K., Molontay, R., & Szabó, M. (2021) A kollégiumi lét egyetemi teljesítményre gyakorolt hatásának vizsgálata. *Statisztikai Szemle*, Vol. 99. pp. 46–79.
- Zhang, W., Zhou, Y., & Yi, B. (2019) An Interpretable Online Learner's Performance Prediction Model Based on Learning Analytics. *Proceedings of the 2019 11th International Conference on Education Technology and Computers*, pp. 148–154.
- Zwick, R., & Himelfarb, I. (2011) The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement*, Vol. 48. No. 2. pp. 101–121.