

# Elsőéves hallgatók pandémia előtti és alatti bemeneti adatainak elemzése modern adattudományi eszközökkel

2023. január 17.

Az oktatás fejlesztése és korszerűsítése mindig magas prioritású feladat volt. A néhány éve kitört koronavírus járvány miatt az oktatás a világ nagy részében távolléti formára állt át. Ebben az környezetben a diákok és hallgatók teljesítménye jól tapasztalható változást mutatott. Ennek fényében tekintettük célszerűnek a Budapesti Műszaki és Gazdaságtudományi Egyetem Vegyészmérnöki és Biomérnöki Karának 2019-es és 2021-es elsőéves hallgatóinak bemeneti adatainak és eredményeinek vizsgálatát adattudományi módszerekkel.

Kutatásunk célja a hallgatók adatai közötti viszonyok feltárása és a hallgatók további eredményeinek előrejelzése ezek alapján, illetve mindezek összehasonlítása a két évben. Egy jól előrejelző modell birtokában a hallgató későbbi teljesítménye a tanulmányai kezdetén megjósolható lenne, így a lemorzsolódásban veszélyeztetett hallgatók számára időben megfelelő segítséget lehetne nyújtani.

Adattudományi eszközök széles spektrumát használtuk a kutatás valamennyi szakaszában. Az adatvizualizációban az oszlop- és szórásdiagramok mellett kiemelhetők a Sankey-féle folyamatábrák, melyeken látványosak az évek közti különbségek. Kísérletet tettünk a hallgatók csoportosítására az adatok alapján, amelyhez háromféle klaszterezési módszert alkalmaztunk. A hallgatók sikerességének előrejelzéséhez szintén számos gépi tanulási algoritmust használtunk. Az első féléves matematika jegyekből képzett jegycsoportokra különböző osztályozó, a félév végi átlagra pedig többféle regressziós modellt építettünk és hasonlítottunk össze.

Az adatvizualizáció során sikerült összefüggéseket találni az adatokban, többek közt kirajzolódott az emelt szintű érettségi pozitív hatása a további eredményekre, illetve látványos különbségeket mutattunk ki az eredmények eloszlásában a két évben. A klaszterezés nem hozott kielégítő eredményt. A prediktív modellezés során a legjobb modellek 80% fölötti hatásfokot értek el az érdemjegyekre nézve, a regresszió esetén pedig a legjobb átlagos négyzetes eltérés 0.4 körül mozgott mindkét évben.

Összességében kijelenthető, hogy a hallgatók eredményessége romlott, de ettől függetlenül továbbra is előrejelezhető. Az oktatás szempontjából érdekes eredmények elérésén túl olyan adattudományi eljárásokat is alkalmaztunk a kutatásunkban, amelyeknek oktatásban történő alkalmazásáról nincs tudomásunk.