

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

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 változást mutatott. Ennek fényében tekintettük célszerűnek az elsőéves Budapesti Műszaki és Gazdaságtudományi Egyetem Vegyész Biomérnöki Kar hallgatói eredményeinek vizsgálatát.

Kutatásunk célja adott bemeneti mutatók és a hallgatók további sikeressége közötti viszonyok feltárása és az utóbbi prediktálása adattudományi módszerekkel. Egy jól előrejelző modell birtokában a hallgató későbbi teljesítménye a tanulmányai kezdetén megjósolható lenne. Így a lemorzsolódásban veszélyeztetett hallgatók számára időben lehetne megfelelő segítséget nyújtani.

Adattudományi módszerek széles spektrumát használtuk a kutatás valamennyi szakaszában. Az adatvizualizációban kiemelhetők a Sankey-féle folyamatábrák, melyeken látványosak az évek közti különbségek. Az adatok klaszterezésénél háromféle módszer lett kipróbálva és összevetve. A legjobb predikciós pontosság érdekében szintén több modellt építettünk fel és hasonlítottunk össze.

Már az adatvizualizáció során sikerült az adatok között némi összefüggéseket és az évek között különbségeket feltárni. Ennek ellenére az adatok nem voltak jól klaszterezhetők. Sikeres lett a prediktív modellezés, a korlátozott adatmennyiség ellenére magas pontossággal jelezték előre modelljeink az osztályokat.