Beadandó feladat

exams\_data­\_writing\_score

Fazekas Márk – EUV0J3

<https://github.com/fzksmrk/adatbanyaszat-alapok>

# A feladat leírása

Egy olyan klasszifikációs modell létrehozása, amivel a „writing test”, azaz az írás teszt eredménye 75% felett lesz-e.

# Az adatok

A diákok demográfiájáról és a tanulásukkal kapcsolatos információkat tartalmazza, mint pl. a rassz, a szülők tanulmányai, részt vettek-e felkészítő órákon.

# A modell

„C5.0” decision tree modell-t választottam, amivel a bináris változóba besorolt, hogy a „writing test score, jeleztem előre.

Az olvasási teszt eredménye a legfontosabb faktor, a tanuló írási teszt eredményének jóslásához. A teljes döntési fa megtalálható a kódban.

Diagram, schematic

Description automatically generated

|  |  |
| --- | --- |
| Table  Description automatically generated | A teszt adaton való validáció során azt láthatjuk, hogy a model működik.  A célosztályunk a „0”. A TP (sikeresen a célosztályba sorolt elemek száma) 123 (200 elem tesztjénél). A TN 52 elem. FN 14 és FP 11 elem.  A 0,875 accuracy jó (ami azt jelenti, hogy 87,5%-át sikerült helyesen bekategorizálni - Accuracy=(TP+TN)/(TP+TN+FP+FN)), és ez az érték ténylegesen jobb, mint a 0,67 NIR (ami a legnagyobb osztály arányát mutatja meg).  A 0,714-es Kappa érték „Jó” ként értelmezhető.  Chart  Description automatically generated with medium confidence  AUC: 0.9251  A 0.9-nél nagyobb AUC esetében a „modell nagyon jó”. |