# Uvod

Promatramo utjecaj odabranih atributa studenata na njihov rezultat na ispitima. Želimo naučiti prognozirati uspjeh studenata na testovima iz pojedinih disciplina, na temelju njihovog spola, rase, dosad stečenog stupnja obrazovanja i slično. Data set koji koristimo je tablica anonimnih studenata u kojem za svakog piše redom: spol, rasa, stupanj obrazovanja, ručak?, priprema, te rezultati iz triju ispita, matematike, čitanja i pisanja. Objavljeno na Kaggleu „Students Performance in Exams“, <https://www.kaggle.com/spscientist/students-performance-in-exams>.

Nešto

# Cilj i hipoteze

Cilj nam je pronaći (naučiti) funkciju koja za dani niz ulaznih podataka za nekog studenta dati dobru predikciju koliki broj bodova bi taj ili takav student mogao ostvariti na ispitu.

Naš skup podataka sadrži informacije o 1000 studenata i studentica. Budući da želimo učiti funkciju koja iz prvih 5 stupaca prognozira preostala tri (ili samo zbroj preostala tri), a ta tri stupca poprimaju samo diskretne vrijednosti, i to prvi 2 (spol), drugi 5 (rasna grupa), treći 5 (stupanj obrazovanja) te četvrti i peti također samo 2 (ručak i priprema), što znači da svih mogućih kombinacija ovih atributa ima konačno mnogo (200) i to manje nego što imamo redaka u tablici (odnosno studenata), znamo da sigurno imamo par studenata (točnije, mora postojati bar jedna petorka studenata) koji imaju sve navedene atribute iste. Međutim, oni vjerojatno neće imati iste bodove na ispitima.

Imajući to u vidu, htjeli bismo da nam funkcija osim nekakvog očekivanog broja bodova (vjerojatno prosjeka za neki još veći skup studenata) vraća i neku ocjenu greške, to jest očekivano odstupanje od očekivanog broja bodova (opet, vjerojatno varijanca). Na primjer, za dvije hipotetske grupe studenata koje bi imale sve atribute iste osim pripreme, očekivali bi bolji (prosječni) rezultat grupe koja je imala pripreme, ali i da većina broja bodova bude bliska tom prosjeku, odnosno da svi imaju otprilike podjednaki broj bodova, a za grupu koja nije imala pripreme, osim što bi imala manji prosjek, očekujemo i veća odstupanja od tog prosjeka, da bude studenata s jako dobrim rezultatom i jako lošim.

# Pregled dosadašnjih istraživanja

Nešto

# Materijali, metodologija i plan istraživanja

Za odabir skupa za učenje i skupa za testiranje željeli bismo osim nasumičnog izbacivanja napraviti i nekakvo „pametno“ izbacivanje. U skup za testiranje stavili bi točno jedan redak iz tablice za svaku od 200 kombinacija atributa. Tako bi između ostalog osigurali i testiranje svih kombinacija koje možda ne bi imali da smo samo nasumično izbacili 200 redaka.